System-wide analysis of absolute protein abundances during the development from spores to multinucleated hyphae of the filamentous fungus *Ashbya gossypii*

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Lars Molzahn

aus

Hamburg, Deutschland

Basel, 2014

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel auf Antrag von

Prof. Dr. Peter Philippsen

Dr. Alexander Schmidt

Dr. Paul Jenö

Basel, den 24. April 2012

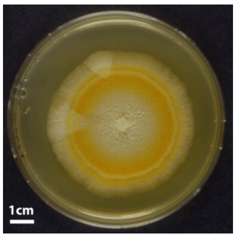
Der Dekan
Prof. Dr. Martin Spiess

Table of contents

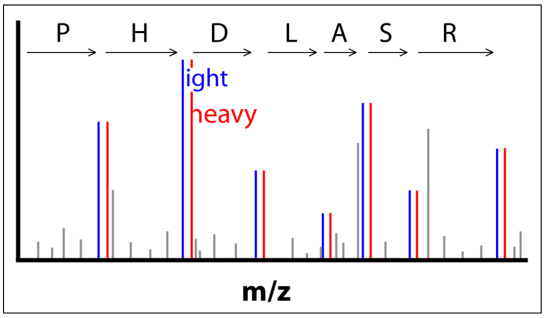
Table of contents

Summary

Summary

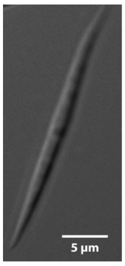


The biological system: *Ashbya gossypii* is a unique organism to study dramatic evolutionary changes within rather short time scales. Phylogenetically, it is closely related to the yeast *Saccharomyces cerevisiae*, which proliferates by budding of uninucleated oval cells. In contrast, *A. gossypii* carrying an almost identical set of genes proliferates by continuous elongation of multinucleated long cells, called hyphae, which frequently branch to form new hyphae but never divide. This generates a fast spreading fungal mycelium, which can cover the surface of an agar plate within a week starting from a single spore. Nutrient limitations in the center of the mycelium induce formation of spores, which can be isolated with high purity. The goal of this PhD thesis was to characterize the proteome at different stages of this rather simple life cycle at a resolution and quantification level so far not achieved with other filamentous fungi.

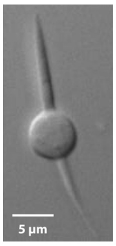


Chapter1 describes the principle of the methods used to analyze peptide mixtures obtained by exhaustive digestions of cell extracts with trypsin. Tandem mass spectrometry (LC–MS/MS) was performed with a reference peptide library extracted from the *A. gossypii* genome. Up to 3,900 proteins of the predicted 4,748 proteins were detected using stringent statistical analyses, and their relative abundances were estimated using two methods. In several cases heavy isotope-labeled synthetic AQUA peptides of known concentration were spiked prior to the MS runs to allow absolute quantifications. Experiments to determine the changing protein compositions and phophorylations during spore germination were performed with biological and technical replicas. For experiments aiming at highest possible resolution the peptide mixtures were prefractionated and each fraction separately analyzed. In total 102 MS runs were performed. The results of most experiments are summarized as tables in the Appendix. Only the high resolution experiments of peptide mixtures (spiked with 40 AQUA peptides) of spores, germ bubbles, and young mycelia are documented and discussed in detail in this PhD thesis.

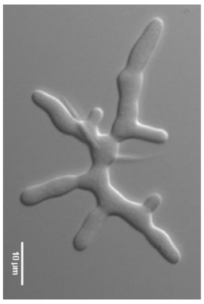
Incubation time	Cell typs	Comments	Ion detection (MS1)	Peptide numbers	Protein numbers	Tables
0h	spores	3 biological + 2 technical replicas each			2144	Table6-1
0h	spores	phospho-enrichment / 2 biol.+2 tech. replicas each	~ 200000	2788	630	Table6-3
0h	spores	absolute quantification (AQUA) / 12 peptide fractions	~454000	116824	3859	Table2-2 / 2-15
2h	pre-germination phase	3 biological + 2 technical replicas each			2145	Table6-1
5h	germ bubbles	3 biological + 2 technical replicas each			2145	Table6-1
5h	germ bubbles	phospho-enrichment / 2 biol.+2 tech. replicas each	~ 200000	2788	630	Table6-3
5h	germ bubbles	absolute quantification (AQUA) / 12 peptide fractions	~454000	121820	3879	Table2-2/3-4/3-6/3-7
7h	mono- / bipolar germlings	3 biological + 2 technical replicas each			2155	Table6-1
9h	bipolar germlings / branches	3 biological + 2 technical replicas each			2155	Table6-1
9h	bipolar germlings / branches	phospho-enrichment / 2 biol.+2 tech. replicas each	~ 200000	2788	630	Table6-3
11h	young mycelia	absolute quantification (AQUA) / 12 peptide fractions	~ 454000	120110	3881	Table2-2 / 4-5
9h+18h	germlings / mature mycelia	protein library / 24 peptide fractions	~ 180000	42589	3265	Table6-2



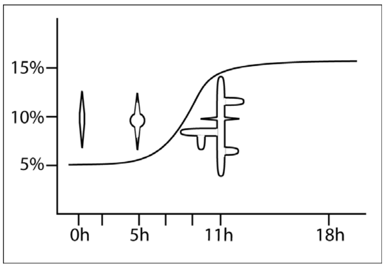
Chapter2 documents in three sub-chapters first the preparations of tryptic peptides from spores and the identification of 3,895 proteins, second the relative abundances of these proteins, and third the approaches used to assign copy numbers per spore to each identified protein. In total, a single *A. gossypii* spore contains $40x10^{6}$ proteins from one copy to several million copies. Special attention was given to the histone proteome including posttranslational modifications. The nucleus in the middle of each spore carries a haploid genome of $9x10^{6}$bp. A genome of this size is associated with approx. 50,000 nucleosomes predicting close to 100,000 copies of each histone per spore. Using AQUA peptides as reference 34,000 copies of histones H3 and H4 were determined confirming that the calculated copy numbers of the other proteins are fairly close to the real copy numbers.



Chapter3 discusses the approach to determine the proteome of isotropically expending germ bubbles. Germination proceeds asynchronously and only mixtures of germ bubbles with different sizes plus morphological unaltered spores can be isolated. Only 5 hours after spore inoculation in liquid medium this mixture consists of 35% germ bubbles and 65% spores. The mixed protein composition was determent and 65% of the known copy numbers in spores were subtracted. The obtained copy numbers for the close to 3,700 proteins most likely reflect the proteome of an average germ bubble, which has not started polar growth yet. In total, $80x10^{6}$ proteins are present in each germ bubble. The adjustment of copy numbers by subtracted of 65% of spore specific proteins resulted in negative numbers in 200 cases strongly indicating that this proteins are actively degraded during the pre-germination phase.



Chapter4 focuses on the proteome of an average young mycelium formed after 11 hours of spore inoculation. Each young mycelium is a giant cell with multiple polar growth sites (hyphal tips) and multiple nuclei. The population of these giant cells is heterogeneous, the number of growing hyphal tips ranges from 3 to 20, the total length of the network of hyphae ranges from 19 to 400 µm, and the number of nuclei ranges from 10 to over 200. We defined the mean of 200 measured mycelia as standard young mycelium, which has seven polarity axes with a total length of 120 µm and contains 48 nuclei. The analyzed cell mixture of young mycelia still contains 15% of spores and germ bubbles the protein compositions of which were neglected because their biomass reflect less the 1% of the total biomass of the growing young mycelia. A standard young mycelium contains $2x10^{9}$ proteins the largest class consisting of ribosomal proteins.



Chapter5 represents the dynamic of the *A. gossypii* proteome during development from spores to young mycelia. Importantly, the contribution of proteins to the total biomass increases from 5% to 15% during this development and remains constant thereafter. Two different types of clustering analyses revealed 800 proteins substantially increased in abundance. Only 200 proteins were degraded during the development and the rest changed marginally or showed up- shifts followed by down-shifts or vice versa. For the septin complex we could show that the relative protein abundances are in the same stoichiometric range and the protein copy numbers have a uniform behavior during development.

General introduction to *Ashbya gossypii*

General introduction to *Ashbya gossypii*

The filamentous fungus *Ashbya gossypii* was identified as causative agent for stigmatomycosis on cotton bolls of the species *Gosssypioieae* (Figure0-1A) and was isolated from tropical and subtropical plant samples (Ashby et al. 1926). Related species of *Ashbya gossypi* are also described in the literature as *Nematospora gossypii* or *Eremothecium ashbii* (Prellinger et al. 1997). The spreading and infection relies on insects of the suborder Heteroptera as vectors of plants pathogens (Mitchell 2004). These insects, called cotton stainers (Figure0-1B), can carry fungal spores in parts of their mouths and help to successfully infect the plant host (Frazer 1944). The internal boll disease leads to a dramatic loss of cotton growing in the affected subtropical regions (Batra 1973), but the use of insecticides is sufficient to control the population of cotton stainers (Dammer et al. 1996). Later, *A. gossypii* was found to naturally overproduce riboflavin (vitamin B2), which leads to yellow colored colonies when grown on agar plates (Sanchez-Marroquin et al. 1970; Demain 1972). Since only minor amounts of riboflavin are essential, it was suggested that the overproduction serves *A. gossypii* spores as a protection by absorbing ultraviolet light (Stahmann et al. 2001). Furthermore, *A. gossypii* strains were developed for industrial production of riboflavin (Stahmann et al. 2000), and strain improvements are still ongoing to use this vitamin for the addition to food products (Sugimoto et al. 2010).

Following the first genetic manipulation of *A. gossypii* (Wright and Philippsen, 1991), the organism became a powerful model organism because of the high efficiency of homologous recombination which is an exception among filamentous fungi (Steiner et al. 1995). Subsequently, further molecular methods were developed and protocols from *Saccharomyces cerevisiae* were adapted and improved for *A. gossypii* (Wendland et al. 2000; Dunkler et al. 2007; Kaufmann 2009). With the complete sequencing of its genome (Dietrich et al. 2004), these techniques became increasingly easier to apply.

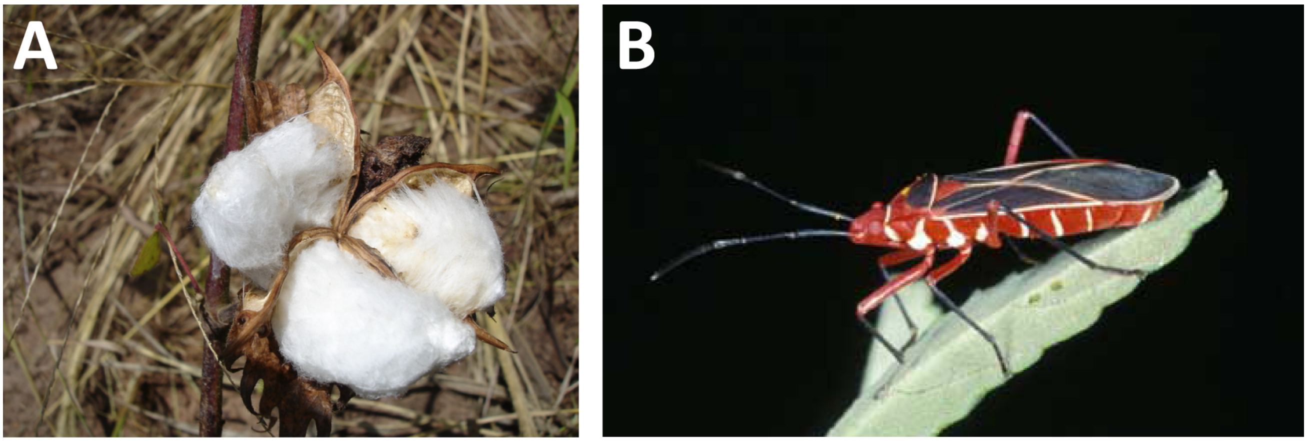


Figure0-1: A) Cotton balls of *Gossypium hirsutum* (Research Institute Senckenberg Botany Department 2008). **B)** Cotton stainer bug (*Dysdercus suturellus*) (Lyle J Buss).

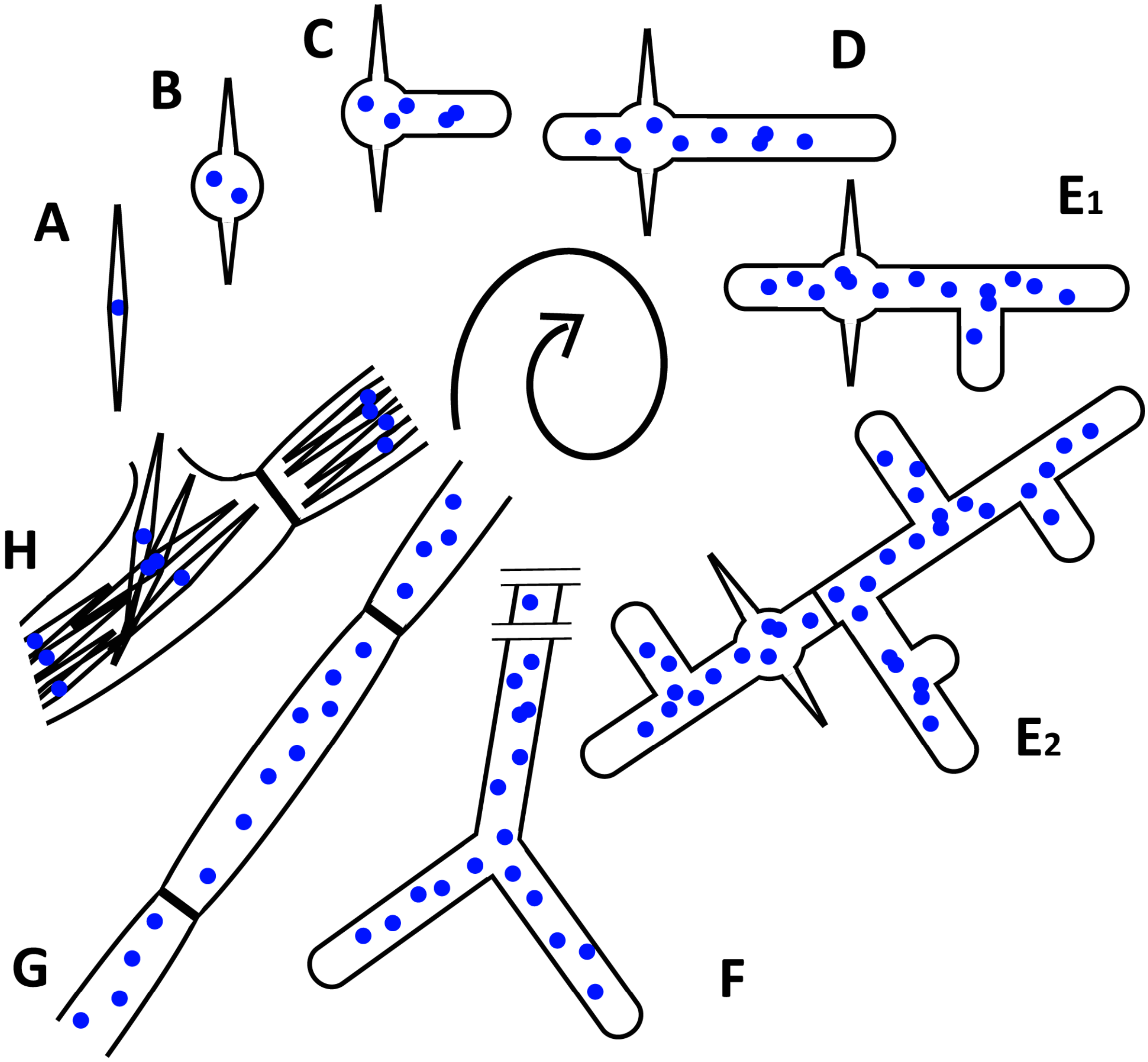


Figure0-2: Life cycle of *A. gossypii*. **A)** Needle shaped-spores. **B)** Germ bubble that is growing in an isotopic manner. **C)** Unipolar germling is defined by emergence of the first germ tube and a switch to sustained polar growth. **D)** Bipolar germling is formed after a second hypha emerges at the cortex of the germ bubble. **E)** Young and advanced mycelium contains many laterally branched hyphae. **F)** Tip splitting occurs at a high growth speeds in mature mycelium. **G)** Early stage sporangia are formed within older parts of the mycelia upon nutrient depletion. **H)** Spores are formed in mature sporangia.

The annotation of the *A. gossypii* genome revealed a high level of genetic similarity between this filamentously growing "yeast" and the budding yeast *S. cerevisiae*. The close relation between both organisms had been proposed in the past (Altmann-Johl et al. 1996; Prillinger et al. 1997). Analysis of both genomes revealed that these two ascomycetes originate from a common ancestor and diverged as individual species about hundred million years ago (Schmitz et al. 2011). During evolution the *S. cerevisiae* lineage underwent a whole genome duplication event thereby increasing the number of chromosomes from eight to sixteen whereas in *A. gossypii* the number of chromosomes were reduced from eight to seven by means of a DNA double strand break in one centromere and fusion of the resulting fragments to the end of two different chromosomes (Dietrich et al. 2004). About 95% of the 4,718 predicted *A. gossypii* open reading frames (ORFs) are orthologues of *S. cerevisiae* genes and 91% are in synteny (Brachat et al. 2003). Almost 500 gene pairs (twin genes) originating from the genome duplication remained in *S. cerevisiae*. Additionally, several million base pair changes and about 300 translocations and inversions separate both organisms (Dietrich et al. 2004; Kellis et al. 2004; Gordon et al. 2009).

The life cycle of *A. gossypii*

Although *A. gossypii* and *S. cerevisiae* share large conserved parts of their genomes, both exert very different life styles. *A. gossypii* exclusively grows as multinucleated long cells, called hyphae, with strictly polar surface expansion at the cell tips. Hyphae also frequently branch but never divide. All nuclei are haploid. On agar plates the growing hyphae form huge networks called mycelia. In contrast, *S. cerevisiae* cells carry a single nucleus, grow by budding, and each mitosis is followed by a cell division. Cell surface expansion of the bud is mainly isotropic, and polar growth occurs only in a short period after bud emergence. Nuclei can either be haploid or dipoid. On agar plates growing *S. cerevisiae* cells form small colonies.

The life cycle of *A. gossypii* as observed in the laboratory begins with needle-shaped spores (Figure0-2A). In the presence of nutrients the spores start to germinate. During the first two to three hours (full medium, 30°C) the spores slowly swell around the central part where the nucleus is localized. During the next four to six hours a germ bubble forms which is the only isotropic growth phase in the life cycle (Figure0-2B). During this process the first nuclear divisions occurs. After two to three nuclear divisions, the germ bubble organizes the first polar growth site perpendicular to the spore axis and a germ tube (the first hypha) is seen after around seven hours (Figure0-2C). This unipolar germling initially grows with an elongation rate of 0.1 µm to 0.2 µm per minute (Knechtle et al. 2003). Within about two hours a second germ tube (the second hypha) develops opposite of the first polar growth site creating a bipolar germling (Figure0-2D). Over time growth speed accelerates, the number of nuclei increases accordingly, and new polar growth sites assemble at the hyphal cortex, resulting in lateral branches that grow orthogonally to the hyphal axis (Figure0-2E). At the branching sites and along the hyphae of young and advanced mycelium, the cytoplasm is partitioned by septa at an average distance of 40 µm (Wendland et al. 2000; Kaufmann et al. 2009). The diameter of hyphae is around 4 µm and slightly increases in faster growing hyphae. About 24 hours after spore incubation the network of hyphae is called mature mycelium. At this stage hyphae switch to a novel branching mode, called tip-splitting, which starts when a critical elongation speed of about 1,5 µm per minute is reached (Knechtle et al. 2003;

Schmitz et al. 2006). The lateral branching pattern is then completely replaced by Y-shaped bifurcations at the tips (Figure0-2F). Fast-growing hyphae can reach a maximum speed of 3,5 µm per minute (~200 µm/h) (Köhli et al. 2008). In mature mycelia, septa have an average distance of 70 µm and the compartments typically contain eight to ten nuclei (Kaufmann et al. 2009). Under nutrient limitation these compartments start swelling and form early stages sporangia in the older regions of the mycelia (Figure0-2G). Eight or twelve needle-shaped spores with a mean length of about 34 µm eventually form and are released from the mature sporangia (figure0-2H). With the release of spores into the environment the life cycle is restarted.

Chapter 1: Technical background

Chapter 1: Technical background

Introduction

Ashbya gossypii carries one of the smallest characterized eukaryotic genomes of 9.2 megabases, which encode 4,748 proteins (Dietrich et al., 2004). The filamentous fungus is an ideal model organism for proteomics studies, since its genome is fully sequenced, well annotated and the number of expressed proteins in a manageable range for extensive mass spectrometry based proteomics analysis (Dietrich et al., 2004). Such a liquid chromatography-mass spectrometry (LC-MS) approach allows the identification of peptide and protein sequences by database searching from acquired tandem mass spectra (Aebersold and Mann 2003) and the quantitative comparison of most expressed proteins across different cellular states. Data interpretation is further facilitated by the fact that the *A. gossypii* genome is closely related to the well-characterized genome of *S. cerevisiae*. In fact, around 91% of the genes are syntenic (Brachat et al., 2003). Consequently, it is very straightforward to predict protein functions in *A. gossypii* based on *S. cerevisiae* homologs that helps interpreting observed protein changes.

Specifically, monitoring relative proteome changes of *A. gossypii* at specific life cycle stages can be used to discover certain protein species involved in cell development. Recently, several studies demonstrated the potential of LC-MS analysis for system-wide protein quantification in various organisms (Washburn et al. 2001, Brunner et al. 2007, Baerenfaller et al. 2008, Schrimpf et al. 2009, Ahrens et al. 2010) and even allow estimation of absolute protein concentrations within cells (Malmstrom et al., 2009).

In contrast to genome and transcriptome analysis, proteome analysis is well suited for gene expression analysis, because it directly determines the actual number of expressed gene molecules in a cell. To date, however, most *A. gossypii* system-wide gene expression analyses only focus on the transcript level by profiling huge numbers of genes in parallel microarray technology (Schena et al., 1996). Transcriptional profiles of filamentous fungi have been reviewed (Breakspear et al. 2007) and also a full expression profiling is monitored for *A. gossypii* (Gattiker et al., 2007). Studies comparing the cellular mRNA and protein abundances have shown poor correlation in general (Maier et al., 2009). For example, the transcript and protein levels in *S. cerevisiae* have a correlation coefficient R^2 of only 0.58 (de Sousa Abreu et al., 2009), underlining the benefits of directly analyzing the proteome of *A. gossypii* by LC/MS. The difference between mRNA and protein levels is a result of the different translational regulation mechanisms during gene expression and variations in protein stability control.

To obtain a global picture of proteome rearrangements during cell growth, the protein profiles and abundances were determined for several growth states and in a system-wide manner. We generated a total of 102 LC-MS runs, acquiring over 2 million tandem mass (MS/MS) spectra. We confidently identifed 4,217 proteins at a 1% false detection rate (FDR) based on the target-decoy search strategy (Elias and Gygi, 2007). Overall, we performed four different sets of quantitative proteomic experiments; (i) a time course of five time points (0h, 2h, 5h, 7h, and 9h of growth), (ii) phosphopeptide enrichment after 0, 5 and 9 hours of growth, (iii) generation of an *A. gossypii* mycelia protein library, (iv) and absolute protein quantification of three time points (0h, 5h, and 11h of growth). Table1-1 summarizing all quantitative proteomic studies using LC-MS techniques, which will be described in more detail below.

Table1-1: *Ashbya gossypii* proteome project.

experiment no.	conditions	runs	specifications	protein count
1	0h	3 biological replicas + 2 technical replicas each	time course	2144
2	0h	12 fractions (generated by OGE)	absolute quantification	3859
3	0h	2 biological replicas + 2 technical replicas each	phospho peptide enrichment	630
4	2h	3 biological replicas + 2 technical replicas each	time course	2145
5	5h	3 biological replicas + 2 technical replicas each	time course	2145
6	5h	12 fractions (generated by OGE)	absolute quantification	3879
7	5h	2 biological replicas + 2 technical replicas each	phospho peptide enrichment	630
8	7h	3 biological replicas + 2 technical replicas each	time course	2155
9	9h	3 biological replicas + 2 technical replicas each	time course	2155
10	9h	2 biological replicas + 2 technical replicas each	phospho peptide enrichment	630
11	11h	12 fractions (generated by OGE)	absolute quantification	3881
12	9h+11h (1:1 mixture)	24 fractions (generated by OGE)	protein library	3265
	total runs	102		

General introduction to mass spectrometry based proteomics

Overview

A large variety of approaches have been developed to accurately quantify Individual proteins within complex biological samples. Most common are Enzyme-Linked Immunosorbent Assay (ELISA) and quantitative liquid chromatography – mass spectrometry (LC-MS). ELISA is based on the interaction of an antibody with the protein of interest. Since the antibody can be linked to a fluorophore or enzyme, the amount of the target protein can be directly determined by fluorescence or electrochemical signals. One major limitation of ELISA is that only a limited number of targeted proteins can be quantified for which suitable antibodies are available. In contrast, mass spectrometry based methods allow unbiased and system-wide quantification of proteins. Such large-scale LC-MS studies follow a general workflow consisting of the following basic steps. First, proteins are extracted from a cell lysate preparation. Second, proteins are denatured, reduced, alkylated and enzymatically cleaved into smaller peptides. Third, sample complexity is reduced by fractionation, for example using OffGel-electrophoresis. After that the peptide samples are separated by high-pressure liquid chromatography and eluted into an electrospray ion source. Finally, peptides are analyzed using tandem mass spectrometry (Aebersold and Mann 2003).

Mass spectrometry basic concept

The mass spectrometry technology is a powerful analytic tool to identify and quantify proteins. Modern LC-MS platforms allow the acquisition of thousands of MS and MS/MS spectra in a single LC-MS run. This results in several thousand peptide and protein identifications (Aebersold et al 2001). Such large-scale proteomics studies are fostered by the development of soft ionization techniques, namely Matrix-assisted laser desorption/ionization (MALDI) (Hillenkamp et al., 1991) and electrospray ionization (ESI) (Fenn et al., 1989). These techniques enable efficient and complete ionization of fragile large biomolecules like peptides, a prerequisite for their MS characterization. Basically, molecules and fragments are measured with the characteristic mass-to-charge (m/z) ratio, which can be translated into the chemical composition and structures of molecules.
All mass spectrometer instruments are composed of three units:

- An ion source; we used electrospray ionization, where ions are transferred from a liquid phase into the gas phase.
- A mass analyzer; which accelerates ions and sort ions by their m/z ratio.
- A detector; which detects ions and defines abundances.

In general, ions are separated by their m/z ratio in the mass analyzer and the signals are processed to generate mass spectra. The ion intensities (signals) are directly correlated with their abundances and therefore can be employed to detect quantitative differences between samples (comparative proteomics). Importantly, ion detectability within a peptide mixture strongly depends on the complexity and purity of the sample as the dynamic range, being the quantitative difference of the lowest and highest abundant ion species in a sample, is limited and depending on the LC-MS platform employed (Domon and Aebersold, 2006). Therefore, very complex peptide mixtures, like whole cell lysates, need to be pre-fractionated to reduce sample complexity. Since each analyte has different ionization efficiencies, the number of detected ions does reflect relative abundances and not absolute numbers. Absolute protein concentrations can be realized via stable isotope dilution methods. We used quantified heavy reference peptides that were spiked with a specific concentration into our samples.

Types of mass spectrometers

In proteomics research different types of mass spectrometers are commonly used for protein identification. The most common are reflector Time-of-flight (TOF), Time-of-flight/Time-of-flight

(TOF/TOF), Quadrupole/Time-of-flight (Q-TOF), Fourier Transform Ion Cyclotron Resonance (FT-ICR), triple Quadrupole (QQQ) and linear ion trap (LIT) mass spectrometers. The reflector Time-of-flight (TOF) accelerates ions, reflects them to compensate the fluctuations, and measures the time for ions to reach the detector, which is mass dependent. TOF/TOF instruments carry a collision cell where selected ions of the first TOF are fragmented and then measured in the second TOF. These specific MS/MS fragment patterns are like a fingerprint for each peptide and can be used to identify its amino acid sequence by database searching (Figure1-3D).

Quadrupole MS instruments measure molecule masses based on the ability to transport ions with a specific m/z ratio via an electric field that changes in a time-dependent manner. Triple Quadrupole selects ions in Q1, in Q2 ions are fragmented, and in Q3 fragments are separated and analyzed. In a linear ion trap, ions are trapped, fragmented, and released via a resonant electric field, which leads to the creation of MS/MS-spectra. The Quadruple TOF is a combination of a triple quadruple instrument consisting of Q1 and Q2 sections fused with a reflector TOF section. In an FT-ICR MS the ions are trapped within a fixed magnetic field, excited by an oscillating electric field and their m/z ratio results from their cyclotron frequency. More details about the advantages and disadvantages of the various mass analyzers for peptide analysis can be found in the review "Mass spectrometry and protein analysis" from Domon and Aebersold (Domon and Aebersold, 2006).

To study protein profiling and absolute protein quantification during the germination of *A. gossypii*, we used a high-resolution linear trap quadrupole (LTQ) - Orbitrap Velos MS-platform, representing a combination of a linear ion trap and an Orbitrap, which represents a new, magnet-free FT-ICR MS and is described below.

LTQ (linear trap quadrupole) Orbitrap Mass Spectrometer

The LTQ Orbitrap Velos consists of the following three main components:

1. A linear ion trap: for ion selection and fragmentation
2. An intermediate storage device (C-trap)
3. An Orbitrap analyzer

Technical details:

Resolution: (apodized) up to 100,000 (FWHM) @ m/z 400
Mass Range: m/z 50 – 2,000 or m/z 200 – 4,000
Mass Accuracy: <5 ppm RMS for 8 h period with external calibration
Dynamic Range: >10,000 between mass spectra, >4,000 between most and least intense mass in one spectrum

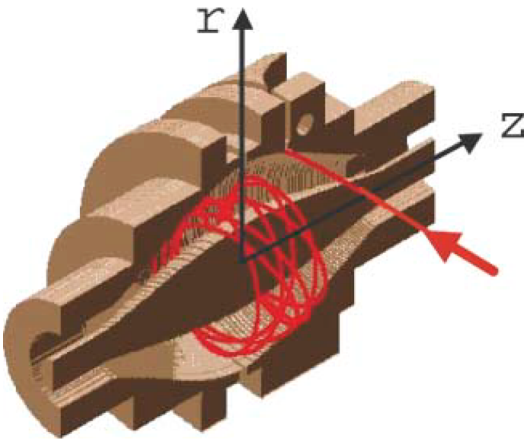


Figure1-1: The Orbitrap mass analyzer. Ions are injected orthogonal to the long axis of the Orbitrap (z-axis) by which the ions is given potential energy in the z-direction (Hu et al. 2005).

The LTQ Orbitrap is a combination of a linear quadrupole ion trap and the Orbitrap mass analyzer. The linear ion trap is composed of three quadrupole sections with a discrete direct current (DC) that contains ions along the axis in the central section of this device (Schwartz et al. 2002). Only the third quadrupole section is functioning as a mass analyzer. The quadrupole mass analyzer filters ions by a time-dependent m/z ratio. The settings of the quadrupole electric field are continuously adjusted, only allowing injected ions of a specific m/z to pass. The Orbitrap mass analyzer was invented by Makarov (Makarov 2000) and is used in combination with the linear quadrupole ion trap. The Orbitrap traps ions in an electric field generated by an outer barrel-like electrode and a coaxial inner spindle-like electrode (Figure1-1). In contrast to the quadrupole ion trap, which is using a dynamic electric field that typically oscillates, the Orbitrap uses a static electronic field (Hu et al. 2005).

Proteome Analysis

Sample preparation (proteolysis)

Proteins have to be specifically cleaved to give a defined mixture of peptides. In proteomics, enzymatic digestion is done via sequencing grade proteases. Different proteases like trypsin, LysC, or GluC are used in proteomic approaches to improve protein identification. The autolytic active serine protease trypsin hydrolyses at the carboxyl side of arginine and lysine residues with a high specificity (Keil-Dlouha et al., 1971), but only poorly cleaves amino acid combinations like RP, KP, RR, RK, KR, or KK (Glatter et al., 2012). This cleavage results in peptides with C-terminally protonated amino acids, providing an advantage in subsequent peptide sequencing (Aebersold and

Mann 2003). Enzymatic digestion of protein mixtures from total cell lysates will generate complex peptide samples that are typically fractionated before MS analysis.

OffGel-electrophoretic fractionation of complex peptide samples

High sample complexity limits the achievable proteome coverage by LC-MS. Therefore we applied isoelectric focusing (OffGel-electrophoresis) (Malmström et al. 2006) to reduce sample complexity to a manageable degree for comprehensive proteome analysis. Figure1-2 shows the schematic pI-based peptide fractionation in liquid phase. The peptides are separated according to their isoelectric point across the pH gradient when an electric current is applied. More specifically, peptides migrate in solution in which the pH has been fixed using ampholytes or an immobilised pH gradient gel (IPG) as soon as an electric field is applied to the pH matching their isoelectric point where peptides are uncharged. The separated peptides are recovered in liquid fractions above the gel strip (Michel, Reymond et al. 2003). We collected samples from three different time points during *A. gossypii* growth (0, 5 and 11h). Results of the experimental OffGel-electrophoresis for the first time point (*A. gossypii* spores) is describe in Chapter II-A (Figure2-4 to 2-6).

LC-MS/MS analysis

In the next step, peptides are separated by high-pressure liquid chromatography (HPLC) using reversed-phase C18 columns. Peptides elute from the stationary-phase by gradually increasing the organic solvent content in an acidic aqueous solution. Since every peptide has a different peptide sequence with different affinities to the hydrophobic stationary phase, they elute at different organic concentrations and therefore at different times during the LC gradient. This results in specific time-dependent base peak chromatograms for each peptide (Figure1-3A). Before the eluted peptides enter the MS, they are ionized *via* nano electro spray ionization (ESI) and directly sprayed under high voltage conditions continuously into the mass spectrometer. During the ionization process, the generated droplets are developing an increasing electric charge density due to solvent evaporation. This leads to Coulomb explosions and finally to a transfer from ions out of the solution into the gas phase before entering the mass spectrometer. During ESI, ions are generated which are differentially charged. These ions of the same peptide species can complicate the exact mass determination.

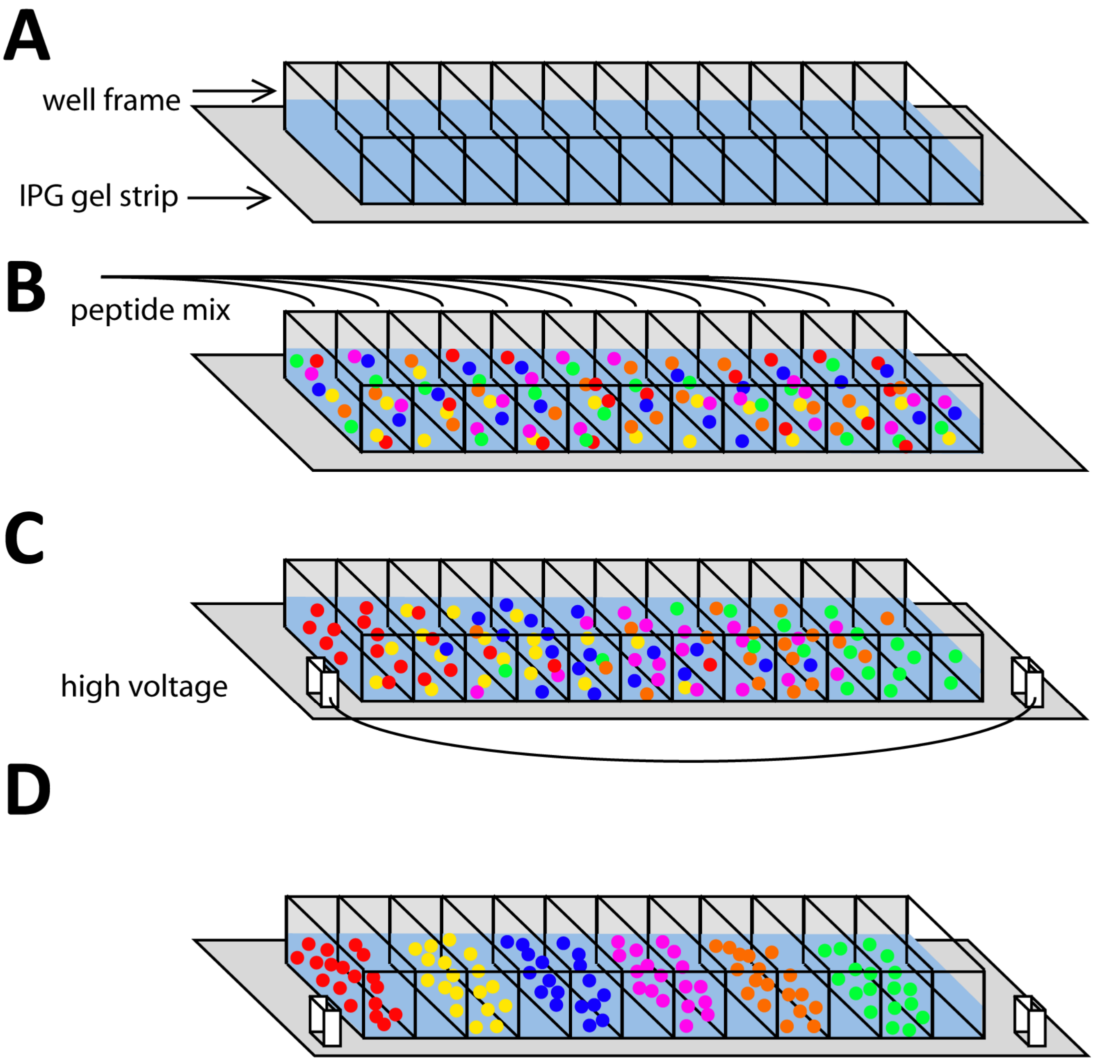


Figure1-2: Flow chart of the peptide separation via OffGel fractionation. **(A)** IPG gel strips are rehydrated and sealed with 12 or 24 well frames. **(B)** The tryptic peptide mixture is distributed in equally amounts per well. **(C)** High voltage moves peptides between the wells towards the pH that matches the pI of the peptide. **(D)** Same peptide species are assembled in the corresponding fractions and collected for the downstream analyses.

Protein identification

To identify proteins by mass spectrometry, two different methods have been developed: Peptide-Mass-Fingerprinting (PMF) (Cottrell, 1994; Pappin, 1997) and tandem mass spectrometry (MS/MS) (Steen and Mann, 2004). PMF identifies proteins based on the precursor ion masses of all generated peptides. In contrast, MS/MS employs peptide characteristics of the product ions, which are generated after peptide fragmentation in a collision cell. The sequence of each peptide is determined from the tandem mass spectrum. Since, PMF only works for low complex samples, like those obtained after one- or two-dimension gel electrophoresis, PMF is not suitable for our large-scale proteomic study. We therefore used tandem mass spectrometry that is explained in more details below.

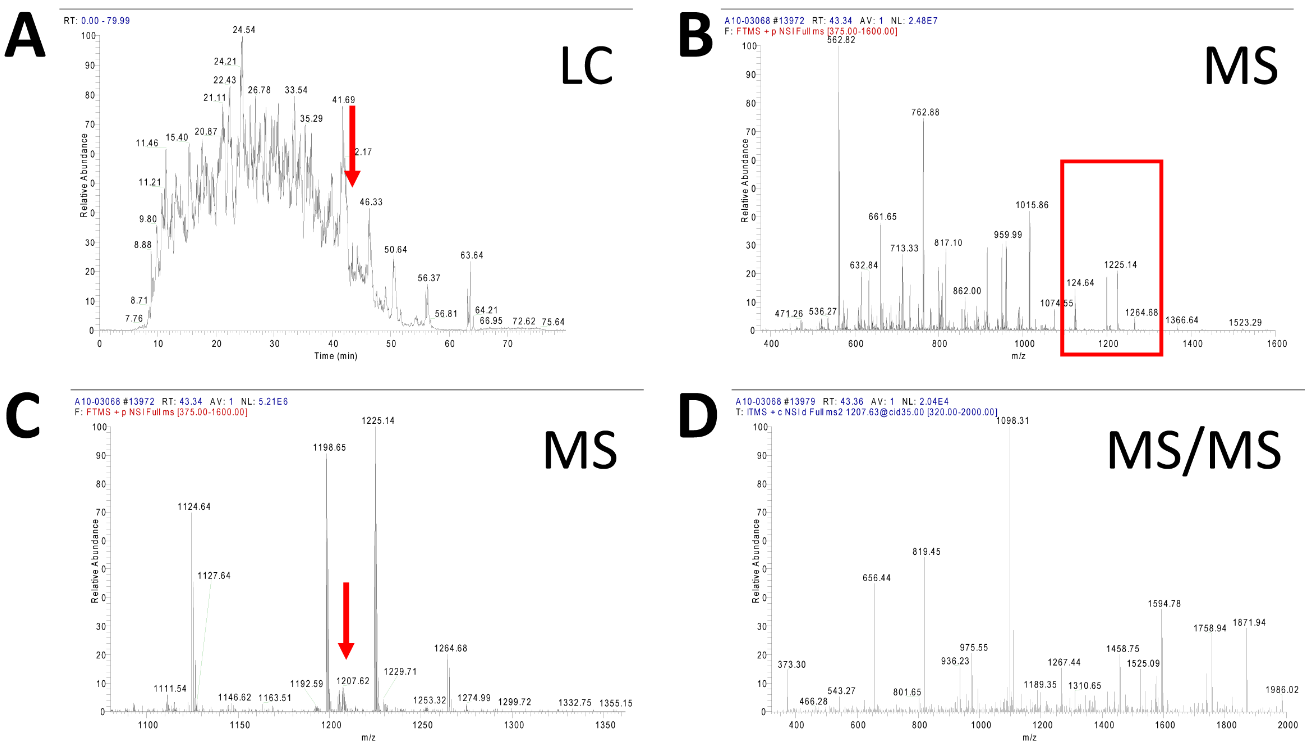


Figure1-3: Flow chart of the mass spectrometry data output visualized by the Xcalibur software. **(A)** Time dependent base peak chromatogram. The time point of the MS 1 scans shown in B is indicated with an arrow. **(B)** MS 1 spectra scaled from 400 m/z to 1600 m/z. **(C)** Zoomed perspective of the MS 1 spectra from 1100 to 1300 dalton. **(D)** MS/MS spectrum of the peptide ion (m/z = 1207.62) marked in C.

Peptide sequencing via MS/MS techniques

Identification of the peptides was done via MS/MS, also named peptide MS-sequencing. This method allows identifying single peptides by their characteristic and unique fragmentation patterns. The selected peptides are separated and transferred into a gas-filled collision cell. The induced dissociation of peptides by colliding with the gas molecules leads to disruption of the peptide bounds. The emerging ions can be described by peptide fragmentation characteristics. The most prominent fragments are called y–ions and b-ions (Roepstorff and Fohlman, 1984), which appear by disrupting the amid bond between the single amino acids at the peptide´s backbone. The developed nomenclature by Roepstorff & Fohlmann is commonly used to annotate MS/MS spectra of peptides (Figure1-4). Depending on the fragmentation orientation, the ions are differently classified by the presence of the C- or N- terminus. Indication of a, b, and c ions demonstrate the presence of the amino group at the N-terminus. If the ions contain the peptides` C-terminus, the ions are titled x-, y-, or z- ions. The subscript classification number defines the count of amino acid residues of the ion. The fragment ions are analyzed and their mass differences corresponding to the remaining residue masses are plotted in MS/MS spectra. An example of a typical MS/MS spectra is shown is Figure1-5. Here, we identified the peptide “AFPINPSDINQLQGVYPSVPEK” of the AEL081W gene product. It is a syntenic homolog of the *S. cerevisiae* mrf1 (YBR026C). For clarity, we only listed the predominant y- and b-ions including the singly charged (y^+, b^+) and the doubly charged ions (y^{+2}, b^{+2}) generated by the doubly protonated peptide ion analyzed. The different fragment masses of the 22 amino acid long peptide are marked in blue for y-ions and red for b-ions. In total, we detected 8 b-ions and 10 y-ions. Due to the specific mass differences between the signals, the amino acid sequence can be calculated. Since protonation occurs more frequently at molecules with a higher pK value, the epsilon-amino group of the C-terminal located lysine is more accessible than the alpha-amino groups of the N-terminal alanine, which leads to more dominant signals of the y-ions. Identification of the single fragments is hindered by the formation of internal fragments resulting from dehydration processes. Furthermore, MS/MS spectra can also be used for *"de novo"* sequencing, but missing and interfering internal fragments limit its application to a few high quality MS/MS spectra.

The total protein sequence coverage of identified peptides varies between different proteins. In principle, all ions with a mass between 700 and 4,500 Daltons are detectable in the MS1 analysis, which covers most peptides consisting of 6 to 40 amino acids. We defined all identifiably peptides of the *A. gossypii* proteome and used the number as an indication of the maximal achievable protein coverage. This proteome catalog contains information of peptides that are specific to one protein and identifiable by mass spectrometry, so-called proteotypic peptides (PTPs) (Mallick et al. 2007). Sequence coverage of two identified proteins is shown in Figure1-6. AFL017W has a low sequence and low peptide coverage (Figure1-6A). Only 4 peptides were detected from a total amount of 21 MS detectable peptides. Of the 241 detectable amino acids only 56 are covered via identified peptides, which is only about 23 percent of the total sequence coverage. In contrast, AAL128C (Figure1-6B) was detected with a higher peptide and sequence coverage. Although only nine peptides were predicted, we could find 16 peptides, including partially cleaved peptides, which results in sequence coverage of more than 70%.

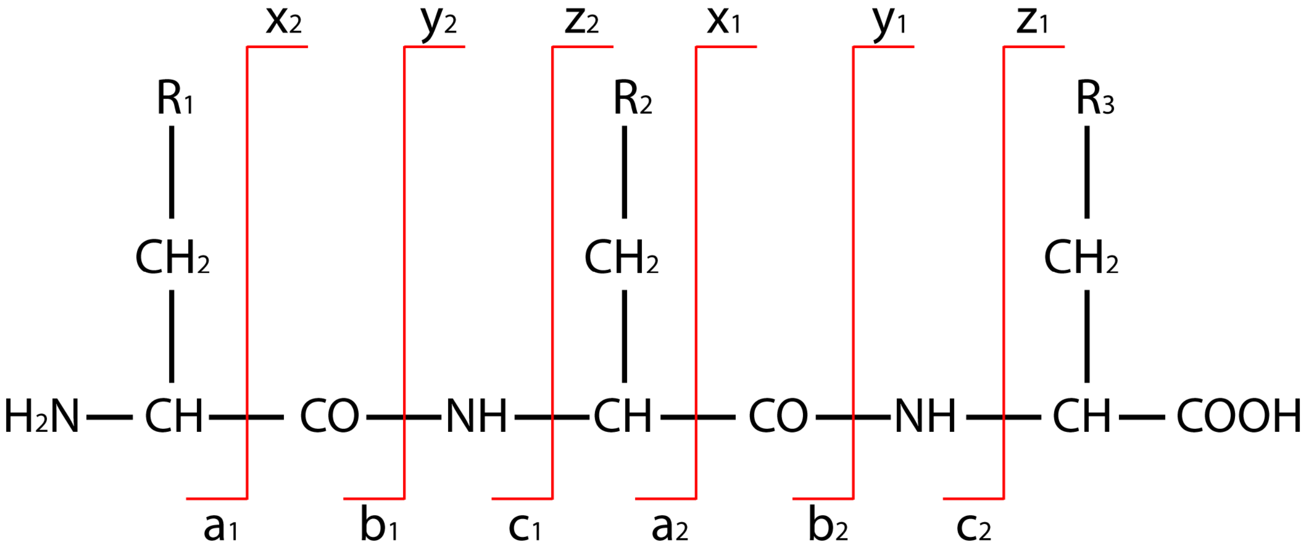
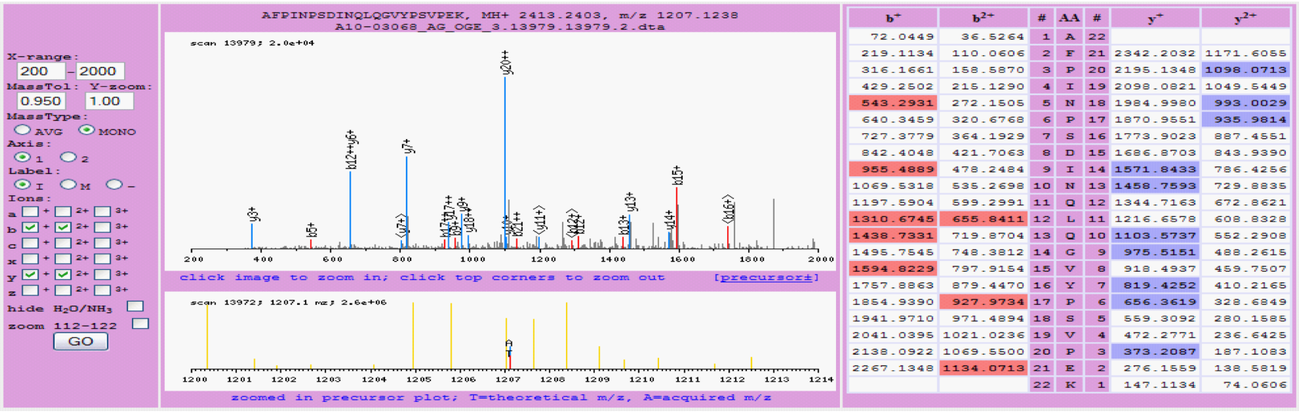


Figure1-4: Schematic fragmentation of peptides. Nomenclature of peptide ions named x, y, z, a, b, and c is based on Roepstorff & Fohlmann.

The total percentage of identified MS/MS spectra is dependent on the quality of both the peptide fragmentation and genome sequence. To visualize the percentage of identified peptides, we plotted the fragmented (blue) and identified (red) peptides using the developed protein sequence library of *A. gossypii* (Figure1-7). In our example, we fractionated the total cell digests into 24 fractions based on the isoelectric point of the cleaved peptides. Overall, we measured around 180,000 MS2 spectra (Figure1-7 stained in blue) and could identify 42,589 *A. gossypii* peptides (Figure1-7 stained in red), representing 3,265 proteins at a false discovery rate of 1%. In total, the identified peptides cover 24 % at the protein sequence database. The low coverage can be explained by missing identification of modified peptides, too small peptides, too large peptides as well as insufficient fragmentation or the simultaneous sequencing of multiple peptides, which complicate unambiguous interpretation of MS/MS spectra. Therefore, in practice, only half of the MS-sequenced peptide ions are generally identified.



AFPINPSDINQLQGVYPSVPEK, MH+ 2413.2403, m/z 1207.1238
A10-03068_AG_OGE_3.13979.13979.2.dta

X-range: 200 – 2000
MassTol: 0.950
Y-zoom: 1.00
MassType: ○ AVG ● MONO
Axis: ● 1 ○ 2
Label: ● I ○ M ○ –
Ions:
a ☐ + ☐ 2+ ☐ 3+
b ☑ + ☑ 2+ ☐ 3+
c ☐ + ☐ 2+ ☐ 3+
x ☐ + ☐ 2+ ☐ 3+
y ☑ + ☑ 2+ ☐ 3+
z ☐ + ☐ 2+ ☐ 3+
hide H_2O/NH_3 ☐
zoom 112-122 ☐

b⁺	b²⁺	#	AA	#	y⁺	y²⁺
72.0449	36.5264	1	A	22		
219.1134	110.0606	2	F	21	2342.2032	1171.6055
316.1661	158.5870	3	P	20	2195.1348	1098.0713
429.2502	215.1290	4	I	19	2098.0821	1049.5449
543.2931	272.1505	5	N	18	1984.9980	993.0029
640.3459	320.6768	6	P	17	1870.9551	935.9814
727.3779	364.1929	7	S	16	1773.9023	887.4551
842.4048	421.7063	8	D	15	1686.8703	843.9390
955.4889	478.2484	9	I	14	1571.8433	786.4256
1069.5318	535.2698	10	N	13	1458.7593	729.8835
1197.5904	599.2991	11	Q	12	1344.7163	672.8621
1310.6745	655.8411	12	L	11	1216.6578	608.8328
1438.7331	719.8704	13	Q	10	1103.5737	552.2908
1495.7545	748.3812	14	G	9	975.5151	488.2615
1594.8229	797.9154	15	V	8	918.4937	459.7507
1757.8863	879.4470	16	Y	7	819.4252	410.2165
1854.9390	927.9734	17	P	6	656.3619	328.6849
1941.9710	971.4894	18	S	5	559.3092	280.1585
2041.0395	1021.0236	19	V	4	472.2771	236.6425
2138.0922	1069.5500	20	P	3	373.2087	187.1083
2267.1348	1134.0713	21	E	2	276.1559	138.5819
		22	K	1	147.1134	74.0606

Figure1-5: Tandem mass spectrum (MS/MS) of the identified peptide AFPINPSDINQLQGVYPSVPEK (1207.1238 m/z) originated from the protein AEL081W. Only y-and b-ions of the fragmented peptide are shown, including the doubly charged ions (y^{+2}, b^{+2}). Peptide fragments, which are and identified, are shown by b-ions are marked in red and y-ions in blue.

To control the error rates in the dataset, a decoy strategy was employed. The additional search for DECOYs is important to determine the number of false positive identifications in the dataset. Since it is not expected to find true matches from the "decoy" database, the counts of detected DECOY peptides directly reflect the number of expected false positive hits that are detected from the normal database (Elias et al., 2005). DECOY sequences have the same precursor mass like the true peptides but the sequences are reversed or randomized. We reserved all *A. gossypii* tryptic peptide sequences within the delectability range of the mass spectrometer. The generated database of DECOYs was merged with the normal *A. gossypii* tryptic peptide database. This combined database ensured identical search parameters for DECOY (false) peptides and true peptides. This can be used to determine the statistical accuracy of the dataset given as false discovery rate (FDR). This FDR can be easily adjusted by using statistical filters to the desired threshold, typically 1%. For all datasets of this study, the mascot peptide score was used to set a peptide identification FDR cutoff of 1%. For protein identifications, the sum of all peptides scores per protein was employed to set the error rate to 1% in all data sets.

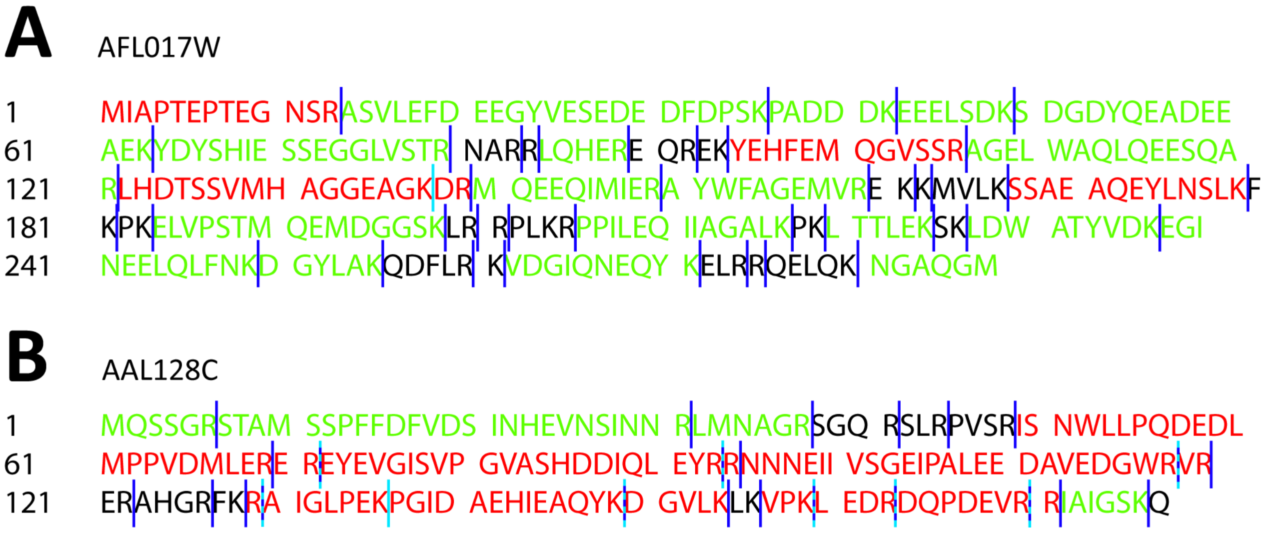


Figure1-6: Theoretical tryptic cleavage pattern of two *A. gossypii* proteins. Dark blue bars indicate all tryptic cleavage sites. Peptides identified by MS/MS are marked with red amino acids. Green colored peptides indicate peptides within the possible detection size, which were not identified. Black marked peptides are not MS-suitable ions. Partially tryptic hydrolyzed peptides are marked with missing cleavage by light blue bars. Light and dark blue dashed lines display right and partial cleavage of different peptides. **(A)** AFL017W is the syntenic homolog of *S. cerevisiae* YBR231C with the common name Aor1. The protein is 287 amino acids long and four out of 21 peptides, which theoretically can be detected, were identified MS/MS. **(B)** AAL128C, the non-syntenic homolog of *S. cerevisiae* Hsp26 is a relative short protein with 179 aa and 9 possible peptides. In total, 16 Hsp26 peptides were detected.

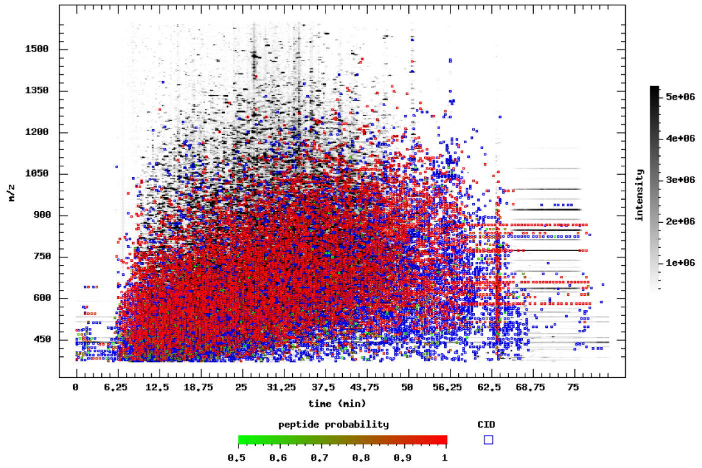


Figure1-7: 2D-plot of MS1 spectra's. The third of 24 fractions of peptide separated by pI via OffGel electrophoresis was analyzed by using the established *A. gossypii* protein library. Acquired MS/MS spectra are indicated in blue and spectra identified from the predicted *A. gossypii* protein database are marked in red.

Label-free quantification

One of the main goals of the analysis of the whole *A. gossypii* proteome is to quantify protein abundances and match them during development stages. Hybrid mass spectrometers like the LTQ Orbitap allow parallel acquisition of MS and MS/MS spectra, thereby greatly increasing duty cycle and MS-sequencing speed. We used label-free quantification to determine expression levels and characterize changes of the different biological life stages. We used the Progenesis LS-MS software (Nonlinear Dynamics) to quantify the protein levels of three life stages using a label-free approach. Additionally, we use an isoelectric separation of the peptides into 12 fractions for each of the three samples collected after 0, 5 and 11 hours of growth. Each fraction of the time course was individually analyzed by LC-MS. The detected ion signals in the MS-scans are plotted across the entire m/z range (Figure1-8A) of one run and fraction. The peak models contain all relevant quantification and positional information for each peptide. In the next step, the same fractions of all three samples were aligned via vector calculation to quantify all detected peaks. We used the 5 hours sample as the reference run, because it is contains most of the proteins of the other two time points (see Figure5-4 for details). For the first fraction, for instance, we calculated in the 11h data set 1,131 vectors and in the 0h data set 887 vectors (Figure1-8B + C). The peak alignment also enables constant peak detection and quantification across all analyzed samples, which can be challenging for other MS quantification techniques (Liu et al. 2004). As an example, a typical alignment of one ion in the three samples is shown in Figure1-9. The peak occurs at the same m/z and retention time positions in all three samples and can be directly compared. The differences in abundances are clearly visibly in the three dimensional projection (Figure1-9C). The ion abundance quantification (based on the integrated peak area) is carried out automatically for all detected ions.

In a next step, peptide sequences are assigned to the detected and quantified peak signals by searching the acquired MS/MS spectra against a sequence database using a search algorithm like MASCOT as described above (Perkins et al., 1999). Figure1-10 shows the aligned peptides, the peak information for the different time points, and the MS/MS spectra used for identification. After exporting the MS/MS data to MASCOT (Figure1-11), the search results are then imported into the quantification software Progenesis LS-MS (Steen and Mann, 2004). The peptide identifications are automatically assigned to its corresponding peak and to its proteins. The abundance of the identified proteins is calculated from the quantitative information of its peptides (sum of intensities) per time point, which can be viewed on the peptide level as well as for the whole protein (Figure1-12) after label-free quantification.

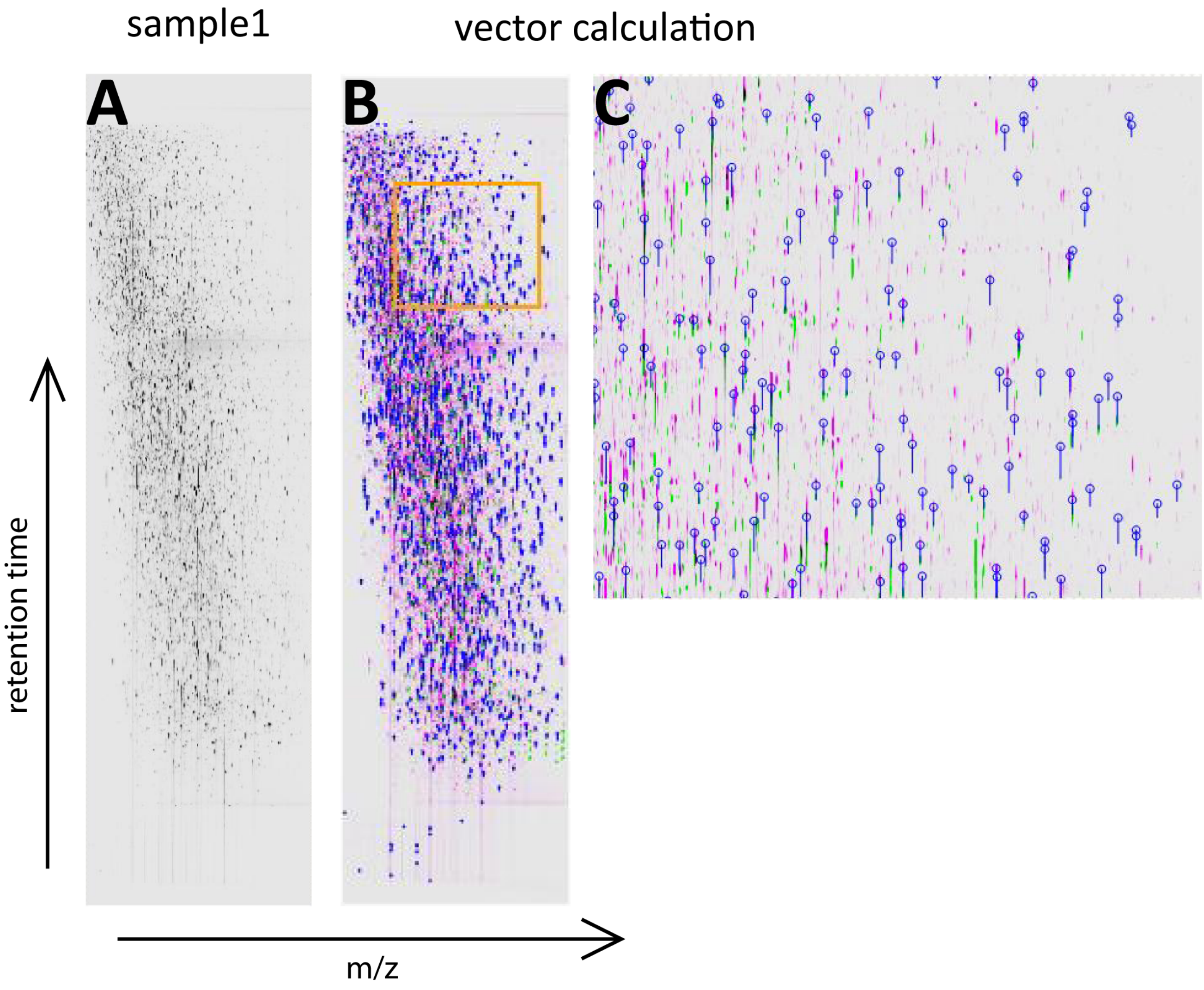


Figure1-8: Peak alignment of same fractions of three different *A. gossypii* samples. Peptides are plotted as m/z versus retention time. **(A)** Peak modeling of one fraction and sample. **(B)** Vector calculation of all three samples and the following alignment. **(C)** Zoomed magnification of the aligned peaks. The identified vectors marked by blue cycles.

The quantitative proteomic data obtained from Progenesis LC-MS data acquisition followed by MASCOT search were statistical analyzed to define significant changes between the three biological samples. We used Spotfire for performing statistical data analyses, Excel for standard calculations and Origin for statistical graphic output of the data. In detail, statistically processed data were generated for all three time points and used for all following data and quantitative analyses. To characterize individual protein fold changes between different samples that are significant, we used statistical testing method by choosing regulated proteins based on the higher value of the 2x standard deviation of each condition (Bantscheff et al., 2007; Dudoit et al., 2004).

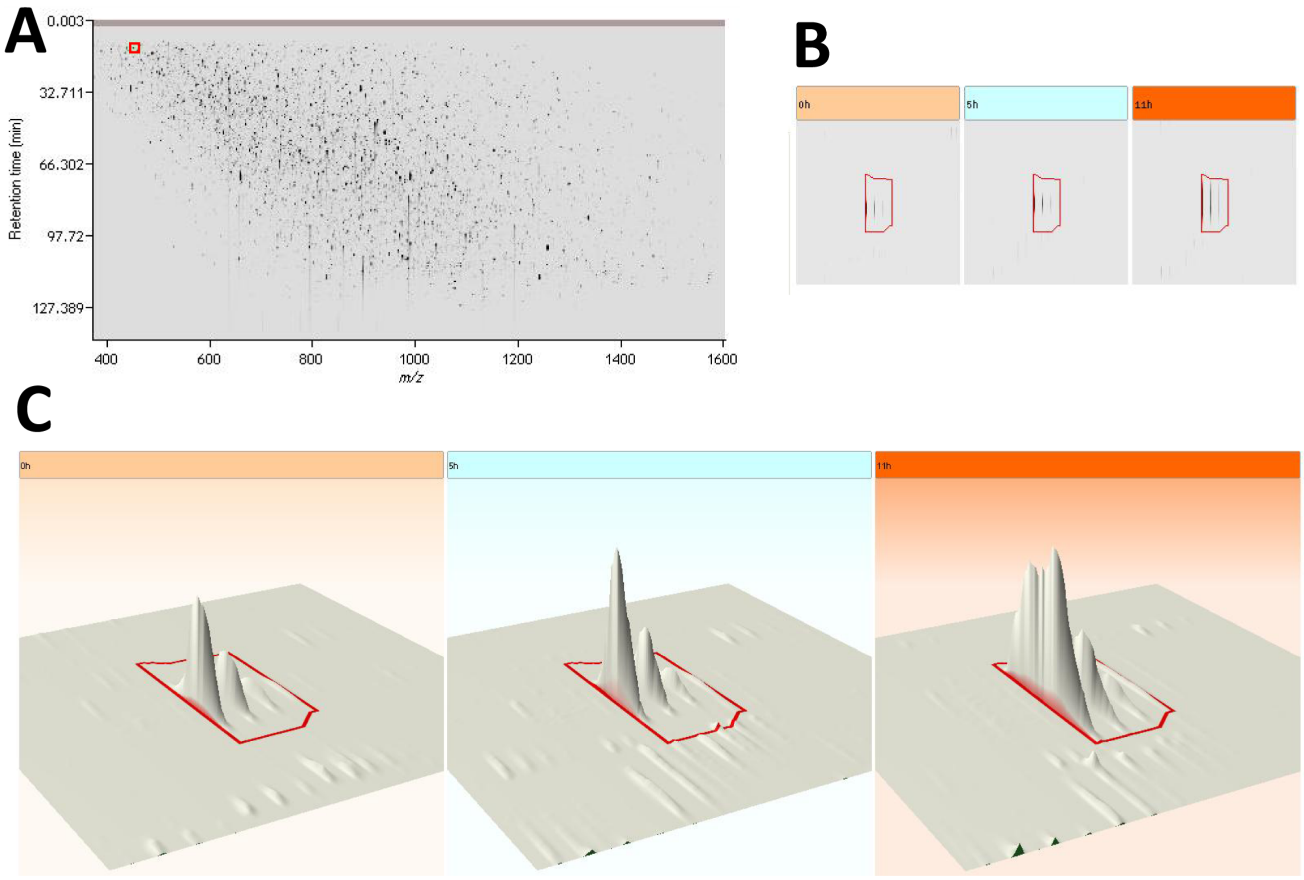
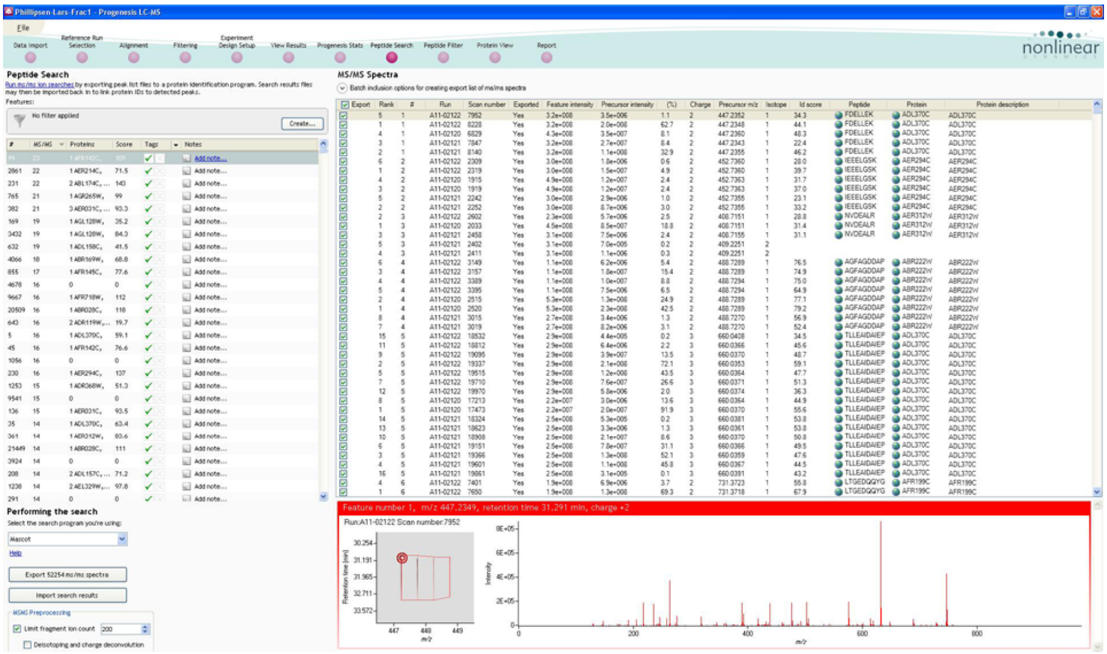


Figure1-9: Peak alignment of peptides across the three samples. **(A)** Alignment of one fraction in plotted by m/z versus retention time. **(B)** Example of an aligned peak across the three time points in a two dimensional graph. **(C)** Three dimensional construction of the quantified peptide in the three biological samples.



Peptide Search

Run ms/ms ion searches by exporting peak list files to a protein identification program. Search results files may then be imported back in to link protein IDs to detected peaks.

Features:

No filter applied

#	MS/MS	Proteins	Score	Tags	Notes
99	23	1 AFR142C,	109	✓ ☒	Add note...
2861	22	1 AER214C,	71.5	✓ ☒	Add note...
231	22	2 ABL174C, ...	143	✓ ☒	Add note...
765	21	1 AGR265W,	99	✓ ☒	Add note...
382	21	3 AER031C, ...	93.3	✓ ☒	Add note...
169	19	1 AGL128W,	35.2	✓ ☒	Add note...
3432	19	1 AGL128W,	84.3	✓ ☒	Add note...
632	19	1 ADL158C,	41.5	✓ ☒	Add note...
4066	18	1 ABR169W,	68.8	✓ ☒	Add note...
855	17	1 AFR145C,	77.6	✓ ☒	Add note...
4678	16	0	0	✓ ☒	Add note...
9667	16	1 AFR718W,	112	✓ ☒	Add note...
20509	16	1 ABR028C,	118	✓ ☒	Add note...
643	16	2 ADR119W, ...	19.7	✓ ☒	Add note...
5	16	1 ADL370C,	59.1	✓ ☒	Add note...
45	16	1 AFR142C,	76.6	✓ ☒	Add note...
1056	16	0	0	✓ ☒	Add note...
230	16	1 AER294C,	137	✓ ☒	Add note...
1253	15	1 ADR368W,	51.3	✓ ☒	Add note...
9541	15	0	0	✓ ☒	Add note...
136	15	1 AER031C,	93.5	✓ ☒	Add note...
35	14	1 ADL370C,	63.4	✓ ☒	Add note...
361	14	1 AER312W,	83.6	✓ ☒	Add note...
21449	14	1 ABR028C,	111	✓ ☒	Add note...
3924	14	0	0	✓ ☒	Add note...
208	14	2 ADL157C, ...	71.2	✓ ☒	Add note...
1238	14	2 AEL329W, ...	97.8	✓ ☒	Add note...
291	14	0	0	✓ ☒	Add note...

Performing the search

Select the search program you're using: Mascot

☑ Limit fragment ion count 200

☐ Deisotoping and charge deconvolution

MS/MS Spectra

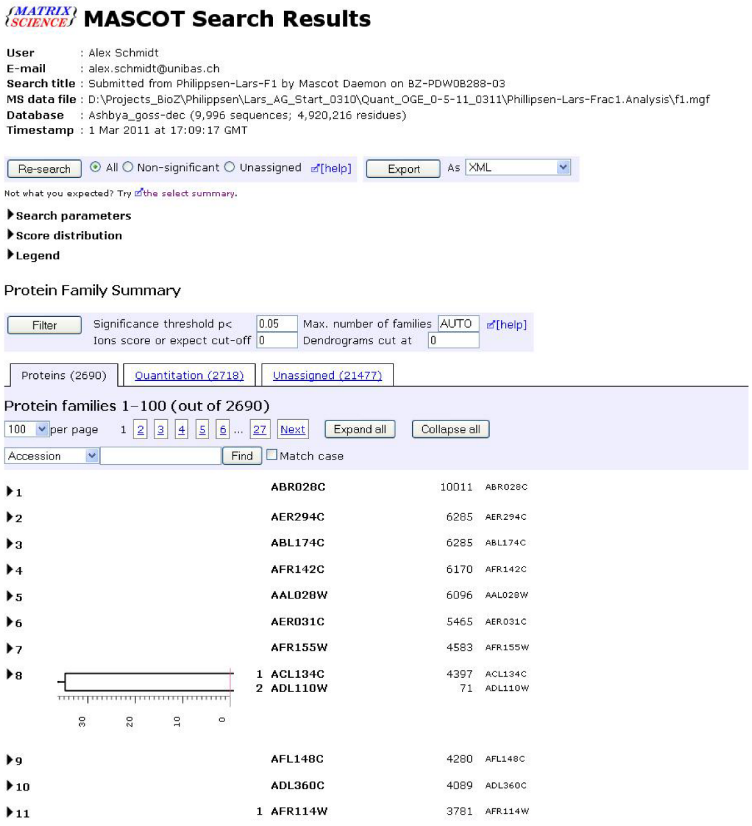
Batch inclusion options for creating export list of ms/ms spectra

☑ Export	Rank	#	Run	Scan number	Exported	Feature intensity	Precursor intensity	(%)	Charge	Precursor m/z	Isotope	Id score	Peptide	Protein	Protein description
☑	5	1	A11-02122	7952	Yes	3.2e+008	3.5e+006	1.1	2	447.2352	1	34.3	FDELLEK	ADL370C	ADL370C
☑	1	1	A11-02122	8228	Yes	3.2e+008	2.0e+008	62.7	2	447.2348	1	44.1	FDELLEK	ADL370C	ADL370C
☑	4	1	A11-02120	6829	Yes	4.3e+008	3.5e+007	8.1	2	447.2360	1	48.3	FDELLEK	ADL370C	ADL370C
☑	3	1	A11-02121	7847	Yes	3.2e+008	2.7e+007	8.4	2	447.2343	1	22.4	FDELLEK	ADL370C	ADL370C
☑	2	1	A11-02121	8140	Yes	3.2e+008	1.1e+008	32.9	2	447.2355	1	46.2	FDELLEK	ADL370C	ADL370C
☑	6	2	A11-02122	2309	Yes	3.0e+008	1.8e+006	0.6	2	452.7360	1	28.0	IEEELGSK	AER294C	AER294C
☑	1	2	A11-02122	2319	Yes	3.0e+008	1.5e+007	4.9	2	452.7360	1	39.7	IEEELGSK	AER294C	AER294C
☑	4	2	A11-02120	1915	Yes	4.9e+008	1.2e+007	2.4	2	452.7363	1	31.7	IEEELGSK	AER294C	AER294C
☑	3	2	A11-02120	1919	Yes	4.9e+008	1.2e+007	2.4	2	452.7363	1	37.0	IEEELGSK	AER294C	AER294C
☑	5	2	A11-02121	2242	Yes	3.0e+008	2.9e+006	1.0	2	452.7355	1	23.1	IEEELGSK	AER294C	AER294C
☑	2	2	A11-02121	2252	Yes	3.0e+008	8.7e+006	3.0	2	452.7355	1	33.2	IEEELGSK	AER294C	AER294C
☑	2	3	A11-02122	2602	Yes	2.3e+008	5.7e+006	2.5	2	408.7151	1	28.8	NVDEALR	AER312W	AER312W
☑	1	3	A11-02120	2033	Yes	4.5e+008	8.5e+007	18.8	2	408.7151	1	31.4	NVDEALR	AER312W	AER312W
☑	3	3	A11-02121	2458	Yes	3.1e+008	7.5e+006	2.4	2	408.7155	1	31.1	NVDEALR	AER312W	AER312W
☑	5	3	A11-02121	2402	Yes	3.1e+008	7.0e+005	0.2	2	409.2251	2				
☑	4	3	A11-02121	2411	Yes	3.1e+008	1.1e+006	0.3	2	409.2251	2				
☑	6	4	A11-02122	3149	Yes	1.1e+008	6.2e+006	5.4	2	488.7289	1	76.5	AGFAGDDAP	ABR222W	ABR222W
☑	3	4	A11-02122	3157	Yes	1.1e+008	1.8e+007	15.4	2	488.7289	1	74.9	AGFAGDDAP	ABR222W	ABR222W
☑	4	4	A11-02122	3389	Yes	1.1e+008	1.0e+007	8.8	2	488.7294	1	75.0	AGFAGDDAP	ABR222W	ABR222W
☑	5	4	A11-02122	3395	Yes	1.1e+008	7.5e+006	6.5	2	488.7294	1	64.9	AGFAGDDAP	ABR222W	ABR222W
☑	2	4	A11-02120	2515	Yes	5.3e+008	1.3e+008	24.9	2	488.7289	1	77.1	AGFAGDDAP	ABR222W	ABR222W
☑	1	4	A11-02120	2520	Yes	5.3e+008	2.3e+008	42.5	2	488.7289	1	79.2	AGFAGDDAP	ABR222W	ABR222W
☑	8	4	A11-02121	3015	Yes	2.7e+008	3.4e+006	1.3	2	488.7270	1	56.9	AGFAGDDAP	ABR222W	ABR222W
☑	7	4	A11-02121	3019	Yes	2.7e+008	8.2e+006	3.1	2	488.7270	1	52.4	AGFAGDDAP	ABR222W	ABR222W
☑	15	5	A11-02122	18532	Yes	2.9e+008	4.4e+005	0.2	3	660.0408	1	34.5	TLLEAIDAIEP	ADL370C	ADL370C
☑	11	5	A11-02122	18812	Yes	2.9e+008	6.4e+006	2.2	3	660.0366	1	45.6	TLLEAIDAIEP	ADL370C	ADL370C
☑	9	5	A11-02122	19095	Yes	2.9e+008	3.9e+007	13.5	3	660.0370	1	48.7	TLLEAIDAIEP	ADL370C	ADL370C
☑	2	5	A11-02122	19337	Yes	2.9e+008	2.1e+008	72.1	3	660.0353	1	59.1	TLLEAIDAIEP	ADL370C	ADL370C
☑	5	5	A11-02122	19515	Yes	2.9e+008	1.2e+008	43.5	3	660.0364	1	47.7	TLLEAIDAIEP	ADL370C	ADL370C
☑	7	5	A11-02122	19710	Yes	2.9e+008	7.6e+007	26.6	3	660.0371	1	51.3	TLLEAIDAIEP	ADL370C	ADL370C
☑	12	5	A11-02122	19970	Yes	2.9e+008	5.8e+006	2.0	3	660.0374	1	36.3	TLLEAIDAIEP	ADL370C	ADL370C
☑	8	5	A11-02120	17213	Yes	2.2e+007	3.0e+006	13.6	3	660.0364	1	44.9	TLLEAIDAIEP	ADL370C	ADL370C
☑	1	5	A11-02120	17473	Yes	2.2e+007	2.0e+007	91.9	3	660.0370	1	55.6	TLLEAIDAIEP	ADL370C	ADL370C
☑	14	5	A11-02121	18324	Yes	2.5e+008	5.3e+005	0.2	3	660.0381	1	53.8	TLLEAIDAIEP	ADL370C	ADL370C
☑	13	5	A11-02121	18623	Yes	2.5e+008	3.3e+006	1.3	3	660.0361	1	53.8	TLLEAIDAIEP	ADL370C	ADL370C
☑	10	5	A11-02121	18908	Yes	2.5e+008	2.1e+007	8.6	3	660.0370	1	50.8	TLLEAIDAIEP	ADL370C	ADL370C
☑	6	5	A11-02121	19151	Yes	2.5e+008	7.8e+007	31.1	3	660.0366	1	49.5	TLLEAIDAIEP	ADL370C	ADL370C
☑	3	5	A11-02121	19366	Yes	2.5e+008	1.3e+008	52.1	3	660.0359	1	47.6	TLLEAIDAIEP	ADL370C	ADL370C
☑	4	5	A11-02121	19601	Yes	2.5e+008	1.1e+008	45.8	3	660.0367	1	44.5	TLLEAIDAIEP	ADL370C	ADL370C
☑	16	5	A11-02121	19861	Yes	2.5e+008	3.1e+005	0.1	3	660.0391	1	43.2	TLLEAIDAIEP	ADL370C	ADL370C
☑	4	6	A11-02122	7401	Yes	1.9e+008	6.9e+006	3.7	2	731.3723	1	55.8	LTGEDQQYG	AFR199C	AFR199C
☑	1	6	A11-02122	7650	Yes	1.9e+008	1.3e+008	69.3	2	731.3718	1	67.9	LTGEDQQYG	AFR199C	AFR199C

Feature number 1, m/z 447.2349, retention time 31.291 min, charge +2

Figure1-10: Peptide sequences are assigned to the MS1 peaks by database searching of the corresponding MS/MS spectra. Results of the matched proteins were provided by the MASCOT search.

All detected peptides and their quantities, including a display of abundance levels were listed in tables. The protein quantification is based on unique peptides only. First, the intensities by matching the same peptide and used the highest peptide confidence score were combined, because most of the peptides were detected with different charge stages. Second, the threshold for the MASCOT ion peptide score has to be adjusted to reach an peptide error rate of 1%. To determine the relative protein abundance from the peptide abundance, commonly two methods are used, the All-method (Schwanhausser et al.) and the Top3-method (Silva et al., 2006). The All-method is defined by computing the sum of all peptide abundances per protein and dividing this result by the number of theoretically detectable peptides, determined by MS detection range. In contrast, the Top3-method focuses only the three most abundant peptides per protein and considers the average of the three peptide as the relative abundance. For the absolute quantification we used selected reaction monitoring (SRM) in combination with stable isotope-labeled peptides, as described in the next section in detail. As statistical reference we calculated the protein score based on the peptide score by summing up the individual peptide scores. In total, we could quantify 3,981 proteins with an error of 1% by using the error DECOY correction (see Chapter2 Figure2-9).



User : Alex Schmidt
E-mail : alex.schmidt@unibas.ch
Search title : Submitted from Philippsen-Lars-F1 by Mascot Daemon on BZ-PDW0B288-03
MS data file : D:\Projects_BioZ\Philippsen\Lars_AG_Start_0310\Quant_OGE_0-5-11_0311\Phillipsen-Lars-Frac1.Analysis\f1.mgf
Database : Ashbya_goss-dec (9,996 sequences; 4,920,216 residues)
Timestamp : 1 Mar 2011 at 17:09:17 GMT

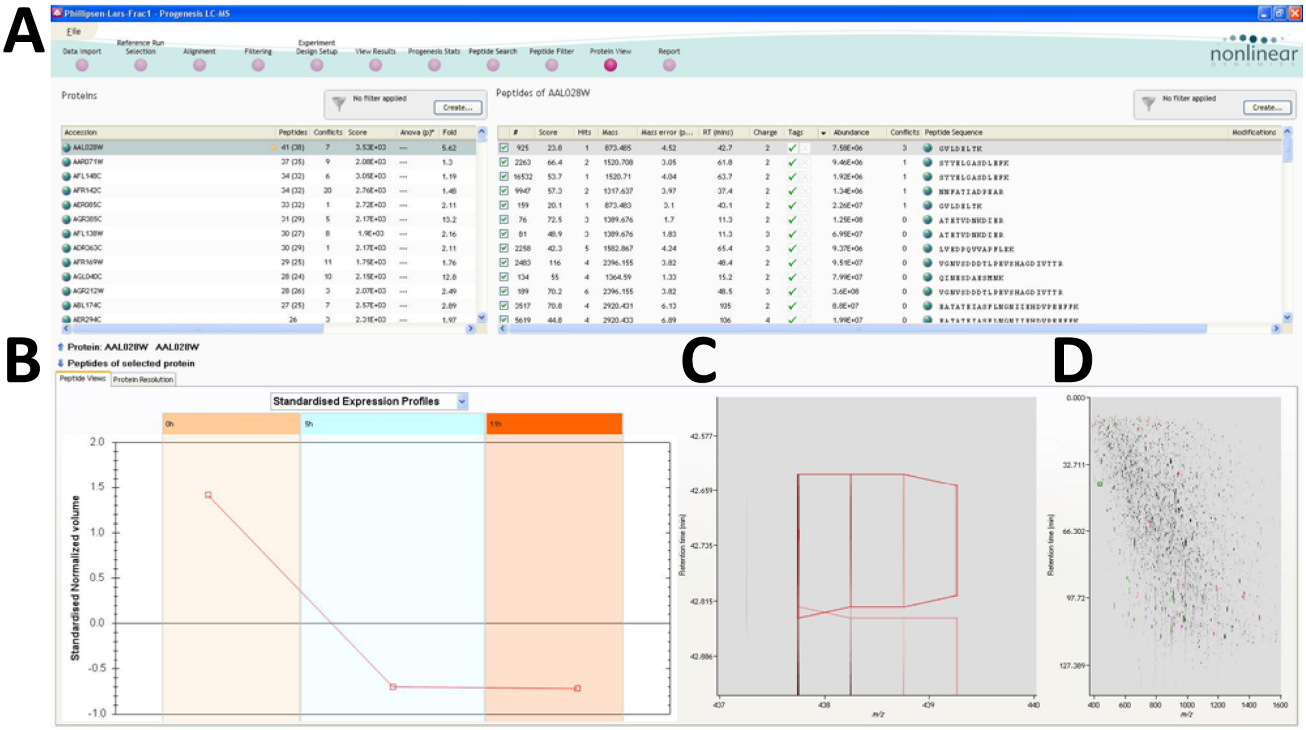
Protein Family Summary

Significance threshold p< 0.05
Max. number of families AUTO
Ions score or expect cut-off 0
Dendrograms cut at 0

Protein families 1–100 (out of 2690)

1		ABR028C	10011	ABR028C
2		AER294C	6285	AER294C
3		ABL174C	6285	ABL174C
4		AFR142C	6170	AFR142C
5		AAL028W	6096	AAL028W
6		AER031C	5465	AER031C
7		AFR155W	4583	AFR155W
8	1	ACL134C	4397	ACL134C
	2	ADL110W	71	ADL110W
9		AFL148C	4280	AFL148C
10		ADL360C	4089	ADL360C
11	1	AFR114W	3781	AFR114W

Figure1-11: Peptides and the corresponding proteins identifications by linking the Progenesis LS-MS data to the external search tool MASCOT.



Accession	Peptides	Conflicts	Score	Anova (p)*	Fold
AAL028W	41 (38)	7	3.53E+03	---	5.62
AAR071W	37 (35)	9	2.08E+03	---	1.3
AFL140C	34 (32)	6	3.09E+03	---	1.19
AFR142C	34 (32)	20	2.76E+03	---	1.48
AER085C	33 (32)	1	2.72E+03	---	2.11
AGR085C	31 (29)	5	2.17E+03	---	13.2
AFL138W	30 (27)	8	1.9E+03	---	2.16
ADR063C	30 (29)	1	2.17E+03	---	2.11
AFR169W	29 (25)	11	1.79E+03	---	1.76
AGL040C	28 (24)	10	2.15E+03	---	12.8
AGR212W	28 (26)	3	2.07E+03	---	2.49
ABL174C	27 (25)	7	2.57E+03	---	2.89
AER294C	26	3	2.31E+03	---	1.97

	#	Score	Hits	Mass	Mass error (p...	RT (mins)	Charge	Tags	Abundance	Conflicts	Peptide Sequence	Modifications
☑	925	23.8	1	873.485	4.52	42.7	2	✓	7.58E+06	3	GVLDELTK	
☑	2263	66.4	2	1520.708	3.05	61.8	2	✓	9.46E+06	1	SYYELGASDLEFK	
☑	16532	53.7	1	1520.71	4.04	63.7	2	✓	1.92E+06	1	SYYELGASDLEFK	
☑	9947	57.3	2	1317.637	3.97	37.4	2	✓	1.34E+06	1	NNFATIADPEAR	
☑	159	20.1	1	873.483	3.1	43.1	2	✓	2.26E+07	1	GVLDELTK	
☑	76	72.5	3	1389.676	1.7	11.3	2	✓	1.25E+08	0	ATETVDNKDIER	
☑	81	48.9	3	1389.676	1.83	11.3	3	✓	6.95E+07	0	ATETVDNKDIER	
☑	2258	42.3	5	1582.867	4.24	65.4	3	✓	9.37E+06	0	LVEDPQVVAPFLEK	
☑	2483	116	4	2396.155	3.82	48.4	2	✓	9.51E+07	0	VGNVSDDDTLPEVSHAGDIVTTR	
☑	134	55	4	1364.59	1.33	15.2	2	✓	7.99E+07	0	QIEESDAESMMK	
☑	189	70.2	6	2396.155	3.82	48.5	3	✓	3.6E+08	0	VGNVSDDDTLPEVSHAGDIVTTR	
☑	3517	70.8	4	2920.431	6.13	105	2	✓	8.8E+07	0	EATATEIASFLMGNIIENDVPEEFFK	
☑	5619	44.8	4	2920.433	6.89	106	4	✓	1.99E+07	0	EATATEIASFLMGNIIENDVPEEFFK	

Figure1-12: Review of all identified proteins and peptides **(A)** and their abundances. Graphic view of the peptide abundance changes during the three time points **(B)**, including the peak alignment of the detected ions **(C)** and their position in the plot based on the m/z and retention time **(D)**.

Absolute protein quantification using selected reaction monitoring

Label-free protein quantification results in relative comparison of proteolytic peptides abundances between different samples (Kuster et al., 2005). Whereas, absolute quantitation can be achieved by spiking in known amounts of isotopically labeled heavy peptides, termed AQUA peptides (Gerber et al., 2003). Isotopic labeling of samples can also be used for accurate relative quantification of protein levels across different conditions (Chakraborty and Regnier, 2002; Gygi et al., 1999; Ross et al., 2004). A large variety of isotopic labeling techniques with their own strengths and limitations have been developed to date (Moritz and Meyer, 2003). This ranges from enzymatic isotope labeling based on digesting proteins in heavy water ($H_2{}^{18}O$) (White et al., 2009) to stable isotope labeling by amino acids in cell culture (SILAC) (Ong et al., 2002) and various heavy labeled chemical reagents that can be covalently attached to protein residues (Ross et al., 2004; Schmidt et al., 2005).

For determining accurate cellular concentrations for proteins of interest, we used absolutely quantified, synthesized isotope labeled peptides (AQUA standard) from Thermo Scientific. In principle, the quantitative MS analysis is based on classic isotope dilution strategies (Streit et al., 2002). Stable isotope-labeled and endogenous peptides are compared by the peak area abundances via tandem MS using multiple reaction monitoring (MRM) (Barnidge et al., 2003). This targeted LC-MS/MS method uses the stable isotope-labeled synthetic peptides as internal standards (Kuhn et al. 2004). MRM has been used to determine proteins with concentrations less than 50 copies per cell in unfractionated lysates (Picotti et al. 2009). For multiple reaction monitoring (MRM), also called selected reaction monitoring (SRM), analysis, a predefined precursor ion and its fragments are selected and monitored over time. We followed the SRM workflow described in the publication "Selected reaction monitoring for quantitative proteomics: a tutorial" (Lange et al. 2008). After selecting peptide candidates for the proteins of interest, one stable isotope-labeled peptide was synthesized for each protein. To generate specific and sensitive SRM assays, only peptides that have a unique sequence in the protein database and that showed high MS signals in the LC-MS analysis were selected (detailed selection criteria see below). For each peptide, the most intense fragments were determined from the MS/MS spectrum and so-called transitions that consist of the peptide precursor and fragment mass were created. These transitions (precursor/fragment ion pairs) in combination with the retention time of the targeted peptides were imported into the Skyline software (MacCoss Lab Software). After collision energy optimizing the transitions of each assay, the specified transitions are used to specifically monitor, via the two mass filters and over time monitored fragments, the predefined precursor ion and its fragments in the complex peptide mixture. We used the TSQ Vantage Triple Stage Quadrupole mass spectrometer for targeted protein quantification experiments.

With its high specificity and sensitivity, SRM allows absolute quantitation of selected proteins within complex peptide mixtures (Barr et al., 1996; Gerber et al., 2003). Usually, the heavy reference peptides are isotopically labeled at the C-terminal amino acids arginine (^{15}N4, ^{13}C6) or lysine (^{15}N2,^{13}C6). Due to the replacements of the abundant natural 12C and 14N isotopes the mass shift of 10 and 8 Daltons, respectively. The selected peptides should give an easy identifiable signature in mass spectra that is unique in the proteome background. We used our established protein and peptide library of *A. gossypii* from the fractionated samples to choose the most suited candidates based on the criteria described recently (Wienkoop and Weckwerth, 2006). Since, several factors affecting the observation of PTPs in SRM, we selected peptides by the following criteria. First, we selected for each targeted protein of interest, peptides which provide the strongest signals. Only unique peptides that are highly detectable were taken. Peptides with partially cleavages or non-tryptic cleavage sites were not considered for SRM. In particular, we avoided peptides with two neighboring tryptic cleavage sites to reduce miss cleavages. In general, we focused on shorter peptides over longer peptides since they generally showed higher MS responses. In addition we tried to avoid and reduce the number of asparagine, methionine, tryptophan, and N-terminal glutamine residues that are prone to artificially induced chemically modifications. For more details see " Selected reaction monitoring for quantitative proteomics: a tutorial" by the R. Aebersold group (Lange et al. 2008).

To balance the concentration of endogenous and its corresponding heavy reference peptide, we split all reference peptides into two groups. Higher/lower reference peptide concentrations were spiked into the samples for the quantification of high/low abundant proteins. AQUA peptide concentrations were chosen to match the median protein concentrations estimated by label-free quantification (top3 method). For all samples, the AQUA peptides were spiked into the samples after digestion and before C18-desalting, fractionation, and LC-MS analysis. Since isotopes have similar chemical and physical properties, the isotopic peptide pairs show identical behavior during all fractionation and separation steps, like OffGel and LC separations. Consequently, co-elution heavy reference peptides and its endogenous counterpart can be analyzed from the same spectra acquired in the mass spectrometer.

To calculate the endogenous protein concentration, we compared the abundance ratios of both, heavy and light peptide (Figure2-18/19). With the known concentration of spiked AQUA peptides we could determine the endogenous peptide concentration (Figure2-20 and Table2-18). For the 40 selected peptides we could precisely quantify the levels for most proteins by using directed MS with an inclusion list (Schmidt et al. 2011). To absolutely quantify all identified proteins, we used the 40 absolute quantified proteins as calibrants to build a regression line and estimate global protein abundances for all identified protein species in the sample. Two recently developed methods were applied and compared. The copy count of protein determined by the AQUA peptide build regression line based on the All-method we termed as intensity based absolute quantification (iBAQ) approach (Schwanhausser et al.). Consequently, we termed the protein copy count determined by the regression line based on the Top3-method, top 3 protein quantification (T3PQ) method (Silva et al., 2006). The list of selected peptides, the absolute quantification of the corresponding endogenous proteins and the global abundance estimation of all identified proteins are presented in chapter2C.

Chapter 1: Proteomics of *A. gossypii* spores

Chapter2 : Proteomics of *A. gossypii* spores

Introduction

To establish developmental changes of the *A. gossypii* proteome we first wanted to identify, if possible, all proteins present in spores. We anticipated that we would find structural spore proteins but also remnants of the spore forming process in addition to those proteins important for the transformation of dormant spores (Figure2-1) to germinating spores with active metabolism and growth. We assumed that the protein composition is very similar in different *A. gosyypii* spores. Spores are forming inside of hyphae in the older parts of the mycelium. Therefore, an essential prerequisite is the isolation of spores free of non-spore proteins. This can be accomplished in *A. gossypii* as described in Chapter2A. Thus, we were confident that all identified proteins originate from spores. Chapter2A also describes the extraction of proteins from highly purified spores, control of the trypsin digestion, pre-fractionation of the peptides by isoelectric focusing and the statistical analyses of the MS/MS runs for each fraction. We next determined the relative abundances of all identified spore proteins and used histones as an example to find post-translational modifications and to compare methods for determine their abundances (Chapter2B). Finally, we employed two methods to estimate the copy number of proteins in a single spore and used 39 synthetic heavy isotope-labeled reference peptides as an independent control of the data (Chapter2C).

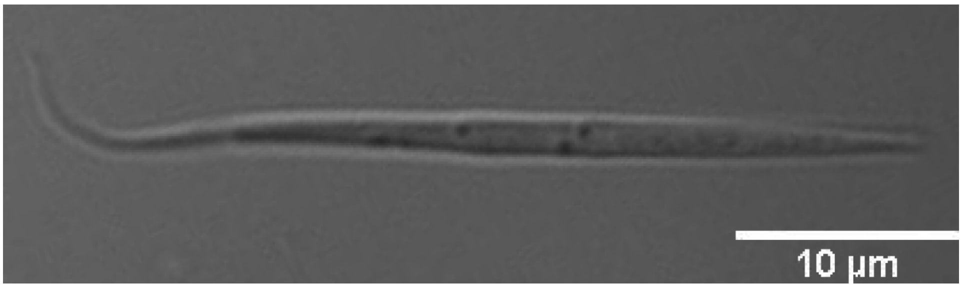


Figure2-1: *A. gossypii* spore, scale bar 10µm.

Part A: Spore preparation and MS spectra analysis

We first analyzed the proteins in three independent spore preparations by standard liquid chromotography tandem mass spectrometry. Up to 2000 proteins could be identified in each preparation (data not shown). In order to increase the number of identifiable spore proteins we decided to employ a high resolution approach with a fourth spore preparation, which is presented in this chapter. The tryptic digest of all proteins was first separated into 12 fractions using isoelectric focusing followed by a 140 minute run of LC-MS for each fraction. Prior to the pre-fractionation 39 isotope-labeled AQUA-peptides of known quantity were spiked into the mixture of tryptic peptides to allow absolute quantifications (see Chapter2C).

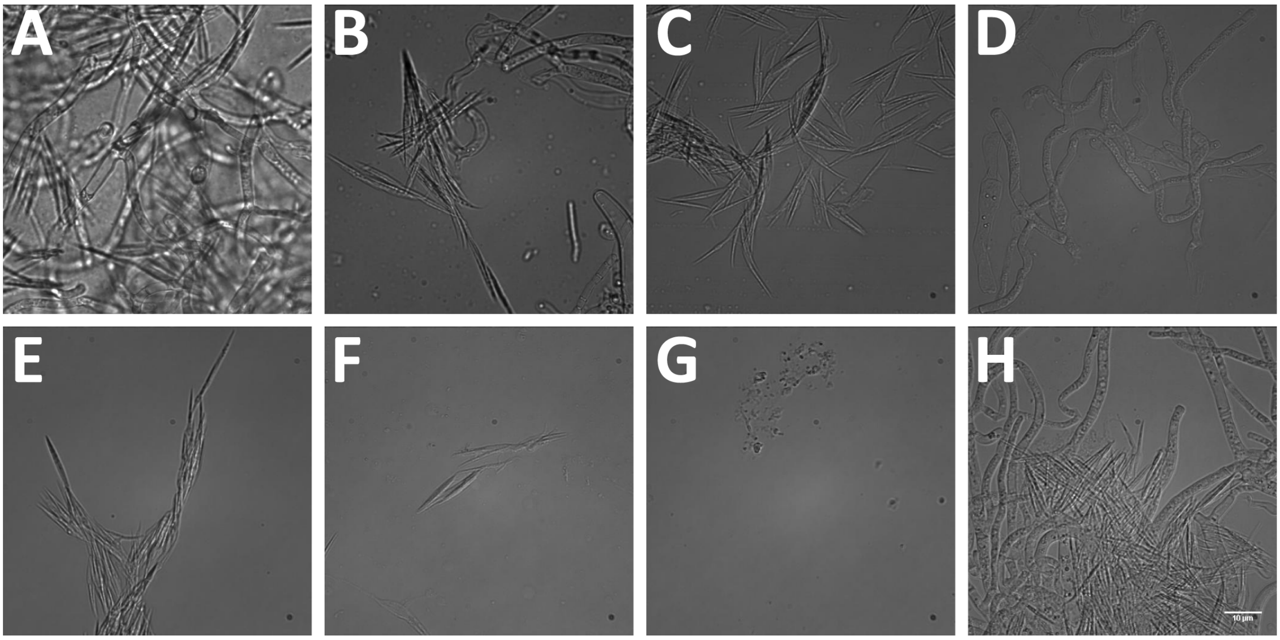


Figure2-2: Isolation and purification of *A. gossypii* spores. Microscopic images of the eight cleaning steps with a 40 times magnification, scale bar is 10µm. **(A)** Six days old mycelium from the center of an AFM plate. **(B)** Spore mixture suspended in water prior to incubation for four hours in a sigmacote coated glass tubes. **(C)** Spores bound to the hydrophobic coated surface. **(D)** Supernatant of the non-bound mycelia particles and spores. **(E)** Spores eluted with 0.1% TritonX after three washing steps. **(F)** Material still bound to the glass surface after the first elution step with TritonX. (G) Control of binding of material to uncoated glass. **(H)** material of the untreated glass surface.

Results

Isolation and purification of *A. gossypii* spores

Ashbya gossypii was grown on AFM agar petri plates to induce sporulation. The middle of the agar was inoculated with *A. gossypii* spores. The radially expanding mycelium covered the agar surface within 6 days. The nutrient depletion led to production of spores which is highest in the center of the colony. To separate spores from cell wall debris and other material we took advantage of the hydrophobic surface of *A. gossypii* spores. By using glass tubes coated with chlorinated organopolysiloxane dissolved in hexane (sigmacote), only spores would bind to the wall of these tubes. Figure2-2 shows images of the isolation and washing steps. First, the center of the mycelium was scraped off, added to bi-distilled water (Figure2-2B) and incubated in the coated glass tubes with constant shaking for several hours at room temperature. The mixture contains *A. gossypii* spores, debris of the lysed mycelium, organelles and released cytosol with high concentration of non-spore proteins. The glass tubes were washed three times with water. Images of the starting mixture, the non-bound material and the bound spores are documented in Figure2-2. The non-bound or only weakly bound material contains mycelia debris, hyphae and spores. In contrast only spores bind to the glass surface. After elution with 0.1% (v/v) TritonX only spores are seen (Figure2-2E). The detergent (TritonX) was used together with mechanical forces (vortex) to release the strongly bound spores from the coated glass surface. A substantial amount of spores stayed attached to the glass surface after the first elution (Figure2-2F). As a control we used the same procedure with non-coated glass tubes. Figure2-2G shows the bound material to the glass surface and Figure2-2H the unbound cell material. Under these conditions neither spores nor other cell material was binding to the glass surface.

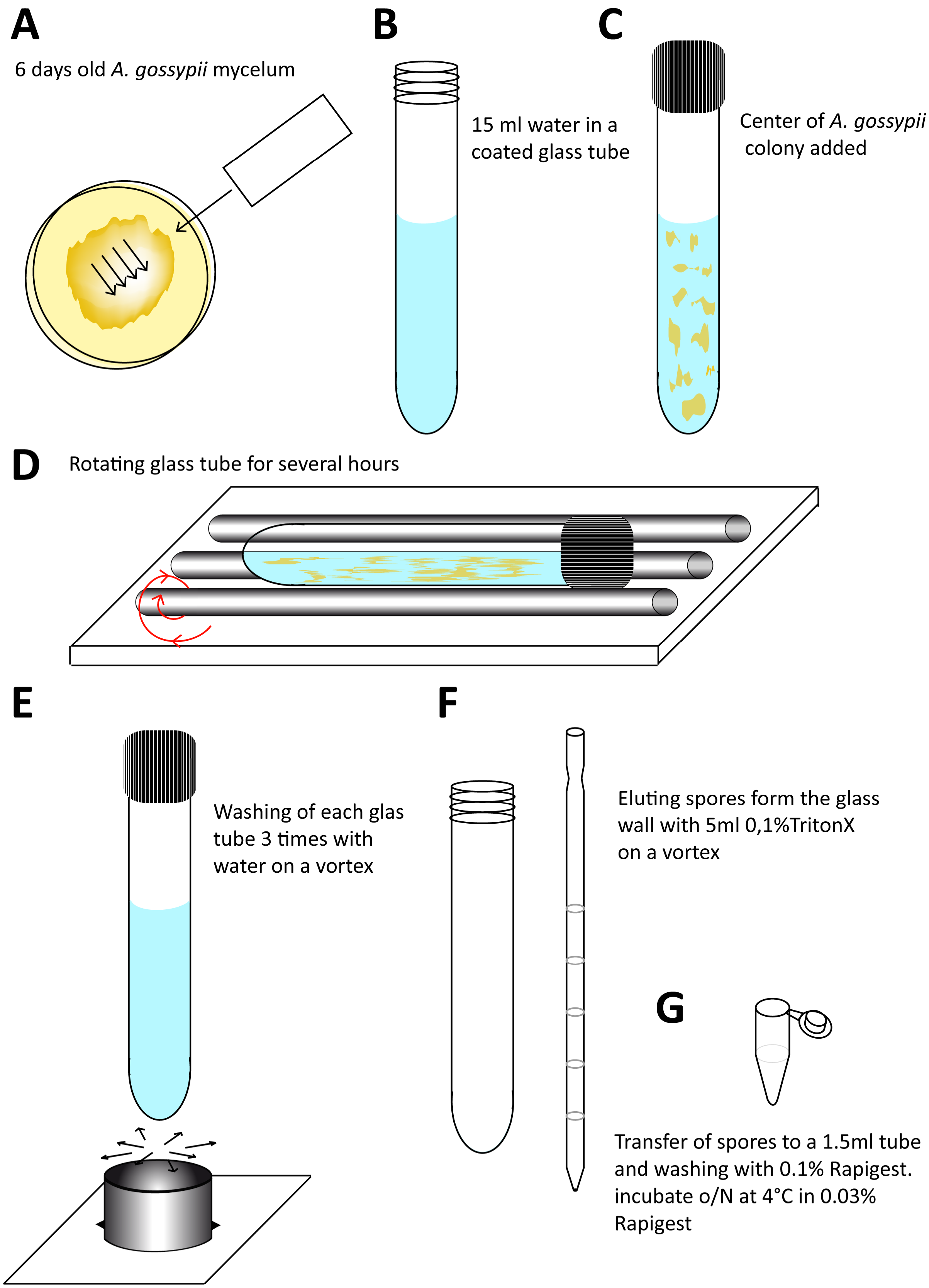


Figure2-3: Flow chart of spore isolation and purification. **(A)** Sketch of a six days old *A. gossypii* mycelium grown on full medium. Glass slides are used to scrape off the central part of the mycelium. **(B)** 25 ml sigmacote coated glass tube with a screw cap. Mycelium was added to 15 ml water **(C)**. Under rotating conditions **(D)** the cell mixture in the glass tube was incubated for several hours. **(E)** Bound spores to the hydrophobic surface were washed with H_2O three times on a vortex. **(F)** Elution of spores with 5 ml 0.1% TritonX under vortexing. **(G)** After washing with 0.1% of the detergent Rapigest we incubated the spores in 0.03% Rapigest overnight.

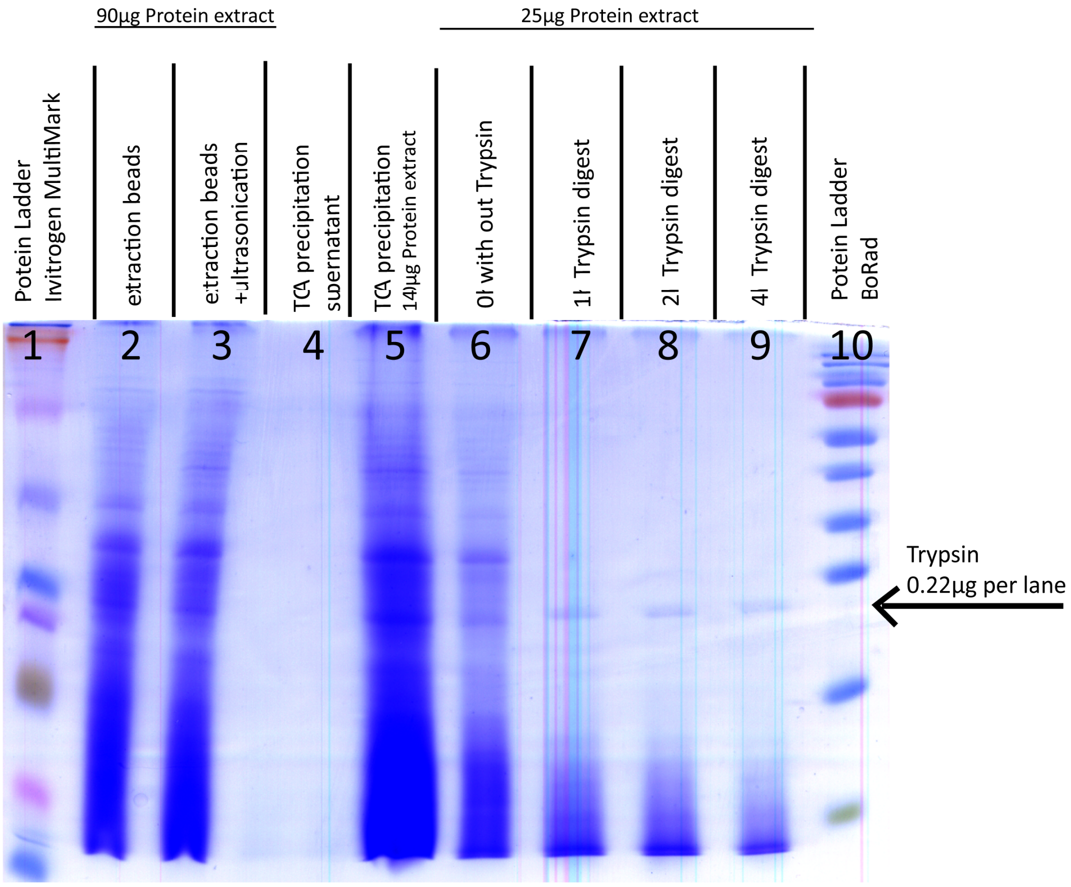


Figure2-4: 15% SDS-PAGE to control spore protein extraction and digestion. Lyses buffer has a concentration of 8M Urea, 0.1% RapiGest, 0,1M Ammonium bicarbonate (ABC). Lane 1: Whole protein extraction from *A. gossypii* spores using silica beads in combination with a fastprep to break the cell wall and lyse the cells. Lane 2: Additional sonification step after the fastprep. Lane 4 and 5: Trichloroacetic acid (TCA) protein precipitation with supernatant. Lanes 6 to 9: Proteolysis with trypsin up to 4 hours. Lane 1 and 10: Protein Ladder.

These experiments were done with screw cap glass tubes as illustrated in Figure2-3. For a single determination of the spore proteome, 40 AFM plates were inoculated with 1 µl wild type *A. gossypii* mycelium and incubated at 30°C for six days (Figure2-3A). Glass tubes were extensively washed, dried and coated with the hydrophobic solution for 4 hours under constant rotation and were dried overnight under a ventilated hood. Next, the coated glass tubes were autoclaved and then used to isolate *A. gossypii* spores. The spore containing mycelium was scratched from the center parts of the 40 colonies and was equally divided into 60 coated glass tubes. Under constant rotating the spore/mycelium mixture was incubated in bi-distilled water which led to the release of most spores from the cell material (Figure2-3D). The suspensions were incubated for three to four hours in the slowly rotating tubes to maximize the coverage of the hydrophobic surface with *A. gossypii* spores. Each glass tube was washed three times with water, applying strong vortex forces to remove remaining cell compartments and organelles bound to the coated glass surface. The spores were released from the glass surface with 5 ml 0.1% TritonX for each tube. During the incubation with TritonX the tubes were vigorously shaken on a vortex to maximize the spore release. Because TritonX is not compatible with the MS instruments, we washed the spores with

RapiGest and stored them overnight at 4°C at a lower concentration. Starting with 40 plates we obtained about 170 mg of clean spores free of other cell material including non-spore proteins.

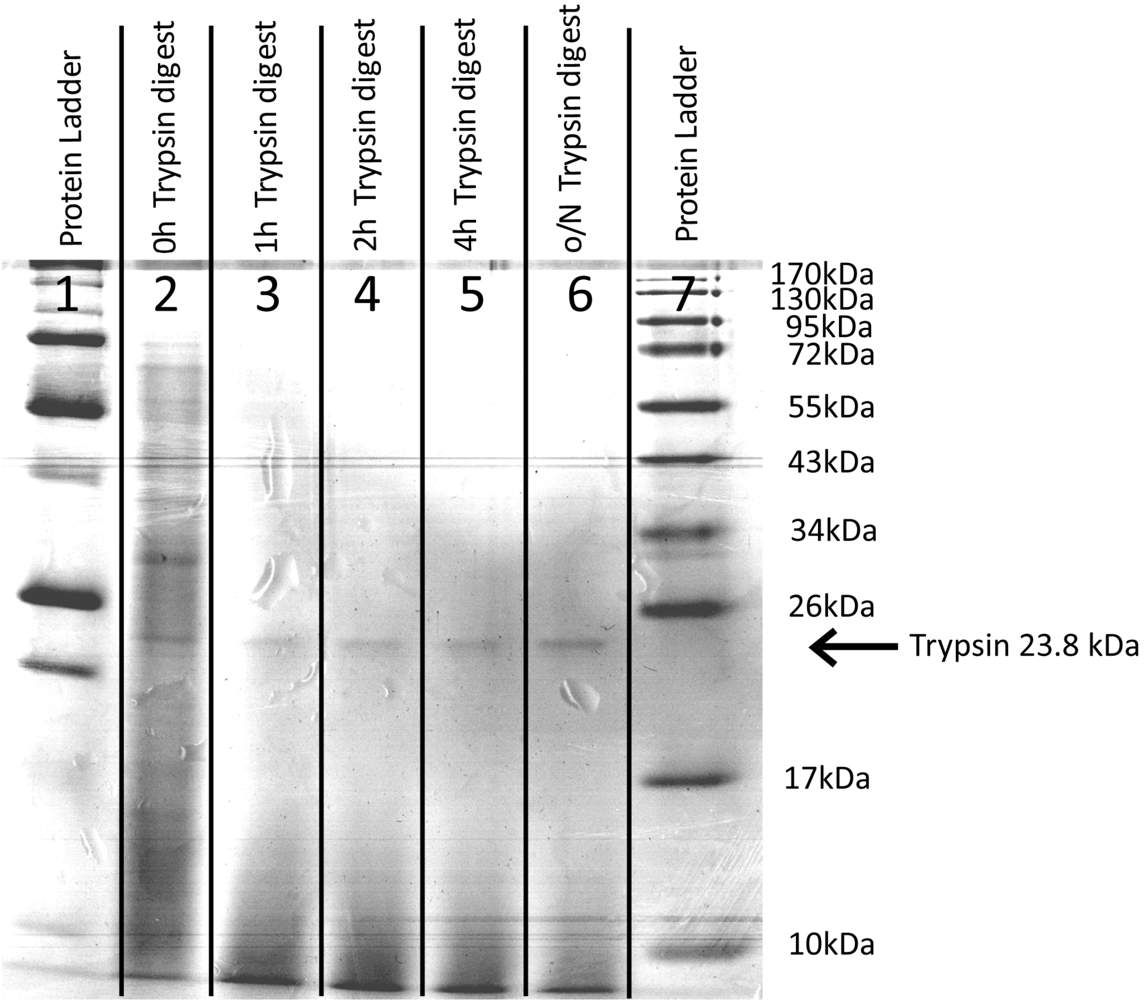


Figure2-5: 15% SDS-PAGE of the tryptic digest of the extracted spore proteins. Lanes 1 and 7: protein size ladder with a range from 170 kDa to 10 kDa. Lanes 2 to 5: samples from the start of trypsin injection (zero hour) and 1 hour to four hours digestion. Lane 6: Additional over night digestion with newly added trypsin.

Protein extraction from of *A. gossypii* spores

After the overnight exposure to RapiGest the fresh *A. gossypii* spores were used for protein extraction. The lysis buffer contained a high concentration of 8M Urea to denature protein and increase their solubility. Additionally, we added 0.1% RapiGest to the 0.1 M ammonium bicarbonate (ABC) buffer for a better solubilization of proteins and to ensure a higher accessibility to cleavage with trypsin. To extract if possible the entire spore proteome we performed six consecutive homogenizations (FastPrep with silica beads) followed by multiple cycles of ultrasound energy. The efficiency of the extraction steps was tested by SDS-PAGE (Figure2-4). Lanes 2 and 3 show a high complexity of the protein extracts of both lysis steps. Next, we precipitated the proteins with trichloroacetic acid (TCA). The redissolved proteins were again tested by SDS-PAGE (Figure2-2 and

Figure2-3). Due to the increased buffer volume of the redissolved proteins the protein concentration as measured with the BCA assay changed from 9.5 mg/ml prior to the precipitation to 5.6 mg/ml. The supernatant was confirmed by SDS-PAGE to contain no or only traces of protein (Figure2-4; lanes 4&5).

Reduction and alkylation of disulfide bonds was done in three steps. First Tris(2-carboxyethyl) phosphine (TCEP), which opens disulfide bridges, was added to 2 mg protein extract (500 µl with 4 mg/ml protein extract). Second, the mixture was incubated with the alkylating agent Iodoacetamide followed by the addition of N-acetyl-cysteine for quenching of Iodoacetamide. After pH adjustment to 8-8.5 trypsin was added to a final concentration of a trypsin protein ratio of 1/50 (for details see materials and methods). Trypsin cleaves after arginine (R) and lysine (K) and the cleavage was tested by SDS-PAGE (Figure2-4; lanes 6 to 9).

Figure2-5 shows that the tryptic digest is essentially complete after four hours of digestion. The tryptic digest of the proteins started directly with the injection of the enzyme. After one hour the majority of the protein is already cleaved after the two amino acids arginine and lysine. Compared to the 4 hours digest stained with Coomassie Brilliant Blue we cannot detect a difference between the lanes in the SDS-PAGE and also not after re-addition of trypsin. The presence of the trypsin band in the protein gel is marked with red in Figure2-5.

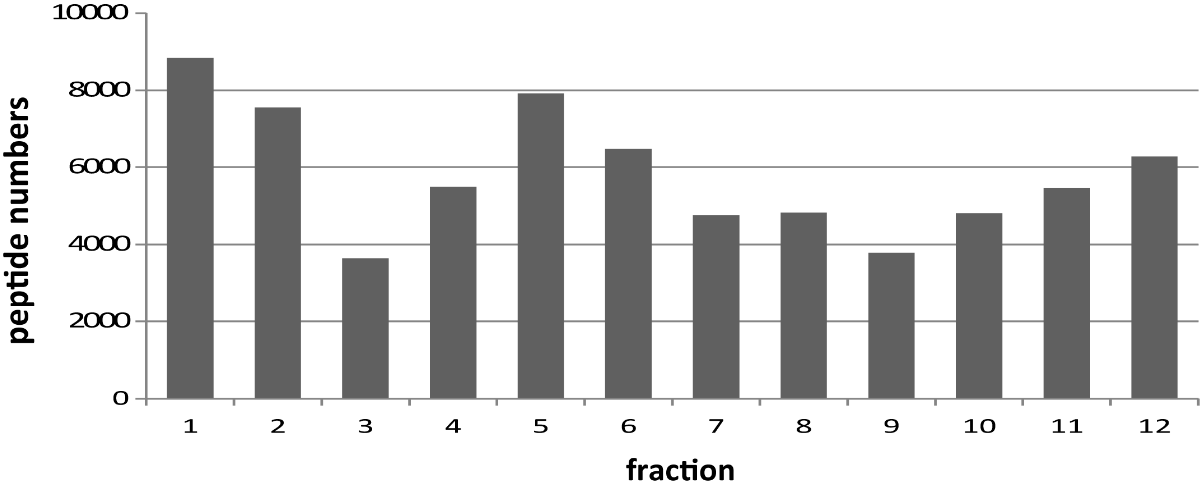
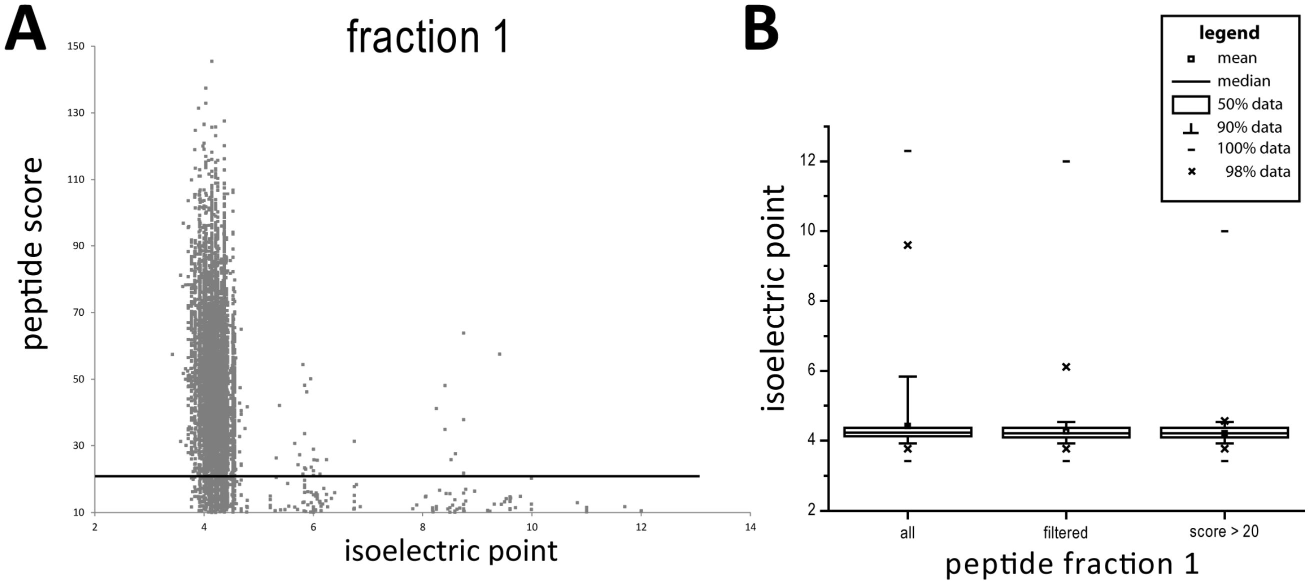


Figure2-6: Fractionation of the tryptic peptides of *A. gossypii* spore proteins. Y-axis shows the peptide numbers. The twelve fractions are plotted on the x-axis, with peptides focusing at acidic pH at the left and basic pH at the right.

Identification of tryptic peptides of the spore proteome

We next desalted the complex peptide mixture with a C18 column. A mixture of 39 AQUA peptides of known concentration was added to the sample prior to the purification step with the column. To achieve the highest possible resolution auf the peptide mixture we first applied OffGel fractionation by isoelectric focusing (pI), which separated the peptides into twelve fractions. Figure2-6 shows the number of unique peptide identifications in each fraction using data from the MS run (see below). Fraction one has the highest number with about 8850 peptides, which have an acid isoelectric point. Only 3,650 peptides were found in fraction three. Peptides with an isoelectric point close to neutral pH accumulated in the fractions five to eight. Fraction twelve is the most alkaline fraction and contains approximately 6,200 peptides. The exact numbers are presented in Table2-1.



C

fraction 1	all peptides	filtered peptides	peptide score > 20
mean (AVERAGE)	4.42	4.26	4.21
median (MEDIAN)	4.23	4.21	4.21
standard deviation (STDEV)	0.94	0.49	0.25
variance (VAR)	0.89	0.24	0.06
n peptide	10203	8836	8146

Figure2-7: Optimization of peptide numbers detected in fraction1 by tandem mass spectrometry. **(A)** Scatter blot of all detected peptides in fraction one. The peptide score is plotted on the y-axis and the theoretically calculated isoelectric point based on the peptide amino acid sequence on the x-axis. A separation line is drawn at a peptide score of 20. **(B)** Box charts of all peptides, filtered peptides and peptides with a score above 20. The isoelectric point is plotted on the y-axis. Borders of 100% of the data are indicated by short lines and 50% of the data are contained in the narrow boxes. **(C)** Table of statistically relevant calculations for fraction 1 that was analyzed in three different ways.

To increase the quality of the peptides that were taken into account to determine the abundance of each protein, we individually analyzed the different fractions and optimized the accuracy for the complex peptide mixture. As an example we reprocessed the analysis of fraction one, described in Figure2-7. All peptides that were identified by the tandem mass spectrometry are plotted according to their peptide score and the calculated isoelectric point (Figure2-7A). The vast majority of peptides is placed in the acidic range of the isoelectric point, but peptides are also detected with a higher isoelectric points. The values of Figure2-7A are compressed into a box chart diagram and plotted at the x-axis in Figure2-7B labeled as "all". With filtering and setting peptide score limitations we decreased the total peptide number but increased the number of correct peptides. The number of all peptides identified in fraction one is 10,203 which changed to 8,146 peptides with a peptide score greater than 20 and acidic isoelectric point. The statistical significance is calculated and described in Table2-1. The mean of fraction one is mainly unchanged with a pH of 4.4. The standard deviation is dropping from 0.94 for all peptides to 0.25 for the filtered and score cutoff peptides. The same is true for the variance that is calculated as only 0.06 for the adjusted values compared to 0.89 for the computed values of all peptides. Both statistical analyses indicate a higher accuracy for the finally accepted peptides. The numbers of peptides in fraction one under unfiltered condition is 2 at pI of 12, 4 at pI of 11, 1 at pI of 10, 20 at pI of 9, 36 at pI of 8, 2 at pI of 7, 41 at pI of 6, 39 at p of 5, 7839 at pI of 4 and 853 at pI of 3. After optimizing the peptide numbers to a score >20 we detected less peptides (2 at pI of 9, 8 at pI of 8, 0 at pI of 7, 8 at pI of 6, 15 at pI of 5, 7292 at pI of 4 and 821 at pI of 3). Clearly the peptides with a high isoelectric point are no longer taken into account and only two peptides with a higher isoelectric point then nine are still present in the data set. Only 33 peptides were recorded with a higher pI value as 4.9.

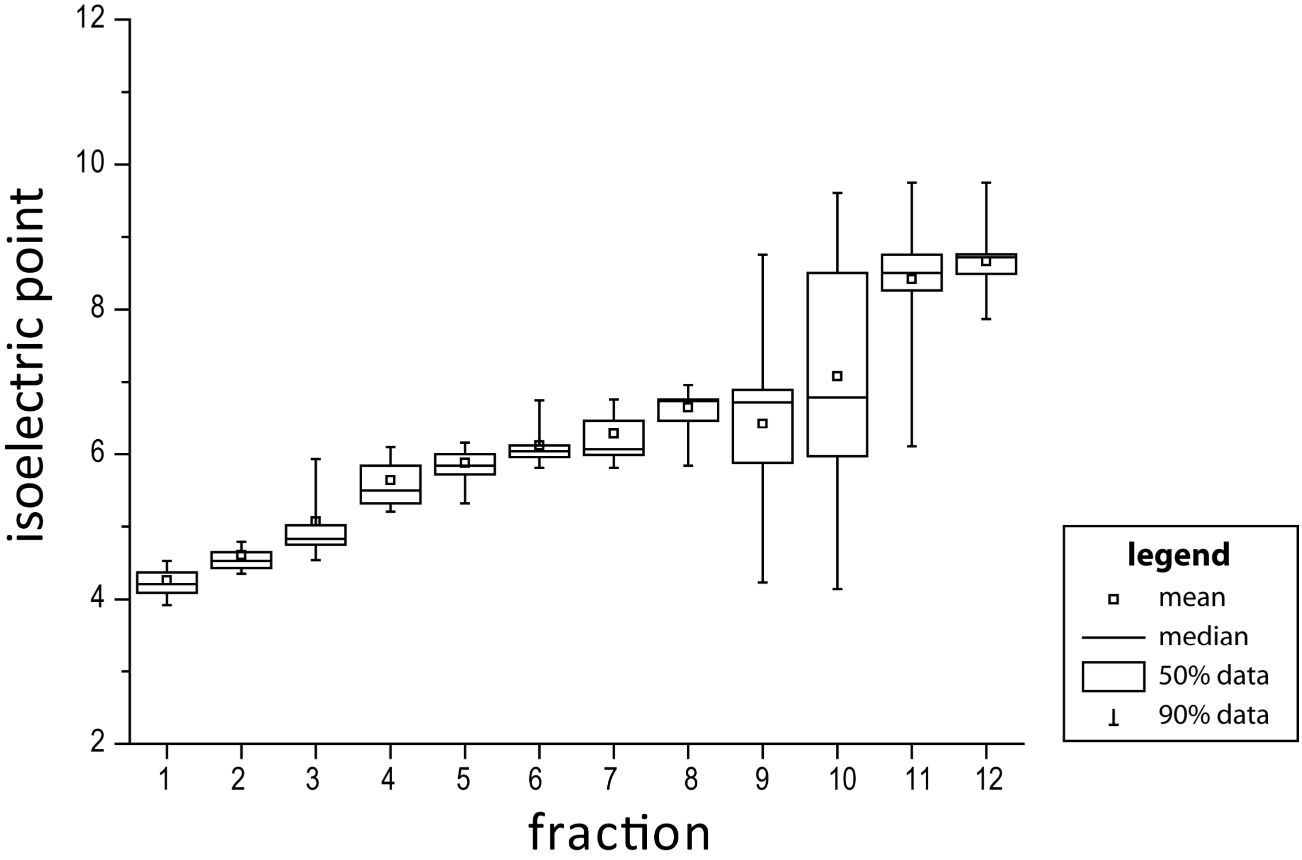


Figure2-8: Isoelectric focusing for all twelve fractions and the theoretically calculated isoelectric points. All filtered peptide fractions (figure2-7) are plotted in the box chart diagrams. 50% of the peptides of each fraction is contained in the boxes. Mean and Median are presented as well as the standard deviation (see insert).

The optimization process described for fraction one was also applied to all fractions (Figure2-8) to gain a peptide list of high quality. The mean for the isoelectric points for each fraction is increasing with increasing fraction numbers. With a pI of 4.26 fraction one has the highest acidic pI. Fraction two is only a little less acidic with a peptide pI average of 4.61. Up to fraction five the peptide pI is still below six. The peptide numbers per fraction are listed in Table2-1 and shows that the amount of acidic peptides is higher than the basic peptides. In total we found 33,500 peptides that have pI below six and only 16,500 with a pI above seven. Fractions nine and ten contain a low number of peptides concomitant with a high distribution of their isoelectric points and a relatively large standard deviation. This is also observed for the other basic peptide fractions eleven and twelve with an average pI of about 8.5. In total the tryptic peptides cover the pI values from 4 to nearly 9.

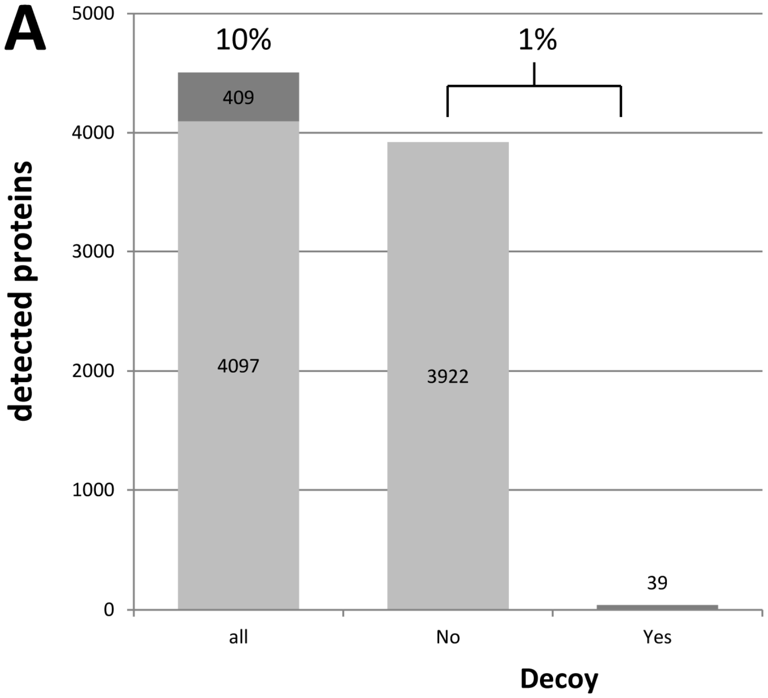
Table2-1: Statistical values for all twelve peptide fractions.

fraction	1	2	3	4	5	6	7	8	9	10	11	12
mean	4.26	4.61	5.07	5.64	5.88	6.12	6.29	6.65	6.43	7.08	8.42	8.66
median	4.21	4.53	4.83	5.50	5.84	6.04	6.07	6.74	6.72	6.79	8.50	8.72
standard deviation	0.49	0.56	0.84	0.73	0.55	0.66	0.77	0.67	1.26	1.68	0.88	0.81
variance	0.24	0.31	0.71	0.53	0.31	0.43	0.59	0.44	1.58	2.81	0.78	0.65
n peptide	8836	7555	3645	5492	7920	6474	4757	4825	3785	4807	5465	6276

LTQ Orbitrap output

Peptides from each of the twelve fractions (Figure2-8) were separated by tandem mass spectrometry to detect and further analyze all charged peptides. The output and the peptide searches converted with Mascot are summarized in Figure2-9B. In total we measured for the spore proteome 454,900 MS/MS spectra. By analyzing these MS2 spectra we could identify 12,8081 peptides in the twelve fractions. By counting all peptides that were identified two or multiple times only once, we collected about 60,000 individual peptides with the MS/MS approach. The absolute threshold of the peptides score was set to 15. This score takes into account the quality of the experimental data and how the experimentally determined peptide sequences match the database sequence. All peptides with a score higher than 15 were used for the protein identifications and for the calculation of the relative abundances. In the next filtering step we deleted all peptides that did not correlate with a theoretical peptide from the tryptic peptide proteome of *A. gossypii*. In total we accepted 52,000 peptides (Table2-1).

After assigning the peptides to proteins by combining the measured peak area of each peptide, we were able to detect 4,500 proteins. To include an error definition we simultaneously ran a search for non-existing inverted peptides called Decoys. Of the 4,506 proteins we identified 409 listed as decoys (Figure2-9A). This means the overall false discovery rate (FDR) is 10%, which states there is a 10% probability that a detected protein is not accurate. Increasing the protein score threshold to equal or greater than 33.66 we lose 155 *A. gossypii* proteins but as well about 400 decoy hits. With an FDR of 1% the detected proteins are statistical significant for true positive matches. For each of the 3,981 *A. gossypii* proteins we computed the abundance with different methods and compared them (see Chapter2B).



B

description	count
MS/MS spectra	454900
found peptides	128081
combined peptides	58902
used peptides [score > 15]	51818
proteins	4506
proteins 1% [score > 33,66]	3922
proteins 1% w/o Decoy	3891

Figure2-9: Mass spectrometry results and computing of the 1% detection error. **(A)** All proteins detected by tandem mass spectrometry and setting to 15 decoy hits. **(B)** Analyzed numbers of spectra and mass translation in spores.

Discussion

To guarantee a high quality and quantity of the spore proteome data we paid attention to the purity of the *A. gossypii* spores and the statistical significance of the experimental data. The washing steps during isolation of spores via hydrophobicity yielded very pure spore preparations. The fractionation of peptides by isoelectric focusing substantially improved the overall resolution. Contaminations with protein and cell particles from the mature mycelium could be essentially excluded. By controlling the fraction of peptides used to compute the protein abundances, the resulting numbers are based on data sets, which have a high quality. For the peptide cut off we used a score of 15, which led to high accuracy (Figure2-7). With the improved data set of detected peptides the protein coverage was close to 85% of the total number of 4,770 proteins predicted by the annotated *A. gossypii* genome sequence (Dietrich et al., 2004; Schmitz and Philippsen, 2011) and updated results. A surprising result is the high number of 3,981 proteins found in spores. This can be explained by the fact that proteins used during the sporulation and proteins needed for efficient germination are stored in spores in addition to the specific spore proteins. For fast germination ribosomes, cytoskeleton proteins, cell surface growth factors, transcription factors, nutrient transporters, functional ATP generation, etc. have to be present in spores (see Chapter2B and C)

Part B: Relative protein abundances in *A. gossypii* spores

Introduction

The previous chapter described the identification of 3,891 proteins in *A. gossypii* spores based on 51,818 tryptic peptides detected by tandem mass spectrometry. Next, we wanted to determine the relative abundance of each protein to reveal the composition of a proteome which enables resting spores to switch, upon contact with nutrients, to the vigorous life of a filamentous fungus. The measured abundances of the tryptic peptide including acetylated and partially cleaved peptides were used to determine protein abundances in the spore proteome using two established methods (Gao et al., 2005; Liu et al., 2004). To validate theses conversions of measured peptide abundances into relative abundances of protein, results obtained for histones were compared with an additional method which solely relies on the most abundant peptide found for each histone. Attempts to determine the absolute numbers of spore proteins in one spore are described in part C of this Chapter.

Results

In the shotgun proteomics experiment, we analyzed the peptides and the fragmentation of the ions by tandem MS (MS/MS) followed by database searching. The proteins were then identified by matching peptide sequences to proteins predicted from the genome sequence taking into account confidence scores of validated peptide identifications. In total, we identified 3,981 proteins that are extractable from purified *A. gossypii* spores, which match 85% of the predicted *A. gossypii* proteins (Dietrich et al., 2004; Gattiker et al., 2007). The relative abundance of each identified protein was determined by two computation methods and the results were documented in a table described in Appendix 1 (Table2-2, complete version). In the first method (All-method), we divided the sum of the abundance of all detected tryptic peptides for a protein by the number of all theoretically detectable tryptic peptides (700Da to 4,500Da). For the second method (Top3-method), we only used the three most abundant peptides of a protein and computed the mean of their peak areas. When only one or two peptides were detected for a protein, the peak areas were divided by only one or two, respectively. To facilitate the comparison of the relative protein abundances, we divided the numbers obtained from both methods by 10,000. This results in an abundance range from 173,000 (Hsp26 at rank 1) to 1 (Urp23 at rank 3,349) and to 0.1 (Spp1 at rank 3,703). For over 200 proteins, abundances below 0.1 were calculated. Because of their very low relative abundance, these proteins are presently not further considered as spore proteins.

Table2-2 short version: Extract of Table2-2 (Appendix 1) summarizing four fractions of spore proteins of very high, high, intermediate, and low abundance.

rank by abundance	accession	common name	0h relative abundance (all *1)	0h relative abundance (top3 *2)	detected peptides	possible peptides	protein size [aa]	protein size [MW]	Score
1	AAL128C	Hsp26	172894	489258	15	9	178	20199	838
2	ADL370C	Tef1, Tef2	51734	203742	66	27	458	49876	3384
3	AGL321W	Sod1	41798	94971	13	7	154	15856	747
4	ADL154C	NOHBY412	41079	175498	20	14	197	21035	1109
5	AFR694W	YLR179C, Tfs1	28817	71915	15	8	204	22333	947
6	AER312W	Tsa1, Tsa2	26675	77194	18	13	197	21486	1076
7	ABL156C	Fmp45	26259	152238	26	24	429	47034	1697
8	ABR068C	Fba1	18909	78537	35	17	360	39319	2287
9	ADL032W	Aco1	18872	192565	73	46	776	84557	4218
10	ABR028C	Cwp1	18731	58675	13	12	763	70107	1138
100	AER015C	Ipp1	2880	13187	28	19	287	32239	1473
101	ABR096C	Ynk1	2856	6912	16	12	151	17080	811
102	ABR146W	Cor1	2841	13414	41	27	445	47536	2372
103	AGL245C	Egd1, Btt1	2811	6931	11	8	161	17357	690
104	ACL148C	Sfa1	2781	15222	25	19	381	40388	1280
105	ACR268C	Mls1, Dal7	2689	12435	41	33	552	62121	2239
106	ABR212C	Met6	2689	21584	67	54	768	85465	3613
107	ABR005C	Rpl14B, Rpl14A	2676	7892	19	12	146	15746	1044
108	ABL105W	Rtn1, Rtn2	2674	13372	17	20	278	29890	945
109	ACL168C	Pfy1	2666	6442	11	8	126	13642	739
1000	AGL013W	Mrps16	103	257	7	8	118	13167	412
1001	AGL061W	Apm3	103	726	14	30	451	49181	764
1002	ABR038C	Vrp1	103	838	10	26	615	61596	421
1003	ADR073W	Rpn5	102	613	23	31	441	51110	1156
1004	AFR062C	Nam9	102	716	32	38	473	53947	1427
1005	AEL253W	Rim2	102	614	10	22	365	40719	588
1006	AGR143W	Oxa1	102	339	16	14	378	41766	717
1007	ABR168W	Pro3	102	398	12	19	283	29550	716
1008	AFL208C	Rpe1	102	419	9	13	239	26181	446
1009	AFR516W	Frs2	102	642	32	35	499	56564	1798
3500	ADR378W	Cin1	0.465	9.91	6	64	1021	117668	265
3501	ACR132W	NOHBY325	0.462	4.62	3	30	513	56479	100
3502	AGR160W	YDR089W	0.460	6.91	3	45	836	94618	103
3503	AAR103C	Ctf4	0.455	7.80	10	56	878	97885	431
3504	AFR416C	YJR061W, Mnn4	0.454	9.85	4	65	1097	128126	124
3505	AFR095C	Dpb11	0.452	5.78	4	41	670	73385	132
3506	AEL250C	Msi1	0.451	3.16	3	21	417	46144	110
3507	AER092W	Vps36	0.426	7.03	2	33	528	59618	88
3508	AGR270W	Med4	0.424	2.40	4	17	294	32113	95
3509	ABL200W	Bni4	0.420	6.11	4	44	789	87530	229

For an overview, four abundance groups extracted from Table2-2 are shown (Table2-2 short version). Ranked by abundances, which were calculated by the All-method, this table comprises the proteins at positions 1 to 10 (very high abundance), 100 to 109 (high abundance), 1,000 to 1,009 (intermediate abundance), and 3,500 to 3,509 (low abundance). The next sections will first comment on the highly abundant proteins in spores and will then focus on a detailed inspection of the histone proteins and their modified peptides.

Highly abundant proteins in spores

Surprisingly, nearly 3,800 different proteins were identified in *A. gossypii* spores. The majority of these proteins were most likely expressed during the sporulation stage but the spore proteome may also include proteins from the mature mycelium co-packaged into the spore compartments (e.g. proteins not completely degraded during sporulation). A substantial part of the proteins very likely represents the set of proteins needed for survival of spores under harsh environmental conditions, and another part may be needed for initiation of germination.

The ranking of the 20 most abundant proteins is listed in Table2-3 and their potential functions are mentioned in Table2-4. The functional descriptions are based on the known function of the homologs in *S. cerevisiae* copied from SGD (*Saccharomyces* Genome Database). The 20 most abundant proteins can be assigned to seven functional groups:

1. Cell wall and membrane (Cwp1 and Om45)
2. Translation (Tef1)
3. Protein assembly and aggregation (Hps26, Cpr1, Tfs1)
4. Carbon metabolism (Tdh3, Fba1, Aco1, Hxk2, Ach1 and Cit1)
5. Redox homeostasis and protein protection against oxidative stress (Sod1, Ypr1, Tsa1)
6. Mitochondria (Fmp45, Pet9 and Ilv5)
7. Cytoskeleton (Act1)

Interestingly, Nohby412 (no homologe in baker's yeast) is one of the highest expressed proteins in spores. It is ranked as number 4 and shows two times higher abundance then Tfs1 the 5th most abundant protein. Recently, this protein was identified as an *A. gossypii* specific cell wall protein (R. Rischatsch and P. De Groot; personal communications). It is remarkable that Table2-3 does not contain ribosomal proteins. The highest abundant ribosomal proteins are Rps29 (rank 25) and Rps8 (rank 30). As shown in Chapters 3 and 4, the relative abundance of ribosomal proteins significantly increases during germination and hyphal growth.

Table2-3: The 20 most abundance proteins in *A. gossypii* spores ranked by the All-Method. (*1) Sum of the abundance of all detected peptides divided by all detectable peptides (All-Method). (*2) Average of the three most abundant peptides (Top3-Method).

no.	accession	common name	relative abundance (all *1) 0h	relative abundance (top3 *2) 0h	detected peptieds	possible peptides	protein size [aa]	protein size [MW]	Score
1	AAL128C	Hsp26	172894	489258	15	9	178	20199	838
2	ADL370C	Tef1, Tef2	51734	203742	66	27	458	49876	3384
3	AGL321W	Sod1	41798	94971	13	7	154	15856	747
4	ADL154C	NOHBY412	41079	175498	20	14	197	21035	1109
5	AFR694W	YLR179C, Tfs1	28817	71915	15	8	204	22333	947
6	AER312W	Tsa1, Tsa2	26675	77194	18	13	197	21486	1076
7	ABL156C	Fmp45	26259	152238	26	24	429	47034	1697
8	ABR068C	Fba1	18909	78537	35	17	360	39319	2287
9	ADL032W	Aco1	18872	192565	73	46	776	84557	4218
10	ABR028C	Cwp1	18731	58675	13	12	763	70107	1138
11	ABR222W	Act1	17046	63127	40	24	376	41688	2355
12	AFR020W	Ach1	16488	104526	51	32	523	57754	2760
13	AGL177C	Cpr1	15950	44902	17	11	162	17384	1077
14	AER428W	Om45	13791	77724	60	32	498	53827	3587
15	ACL198W	Ilv5	12315	80406	48	29	394	43687	2620
16	AER401W	Ypr1, Gcy1	12314	42113	43	22	313	34774	2512
17	AAR004C	Cit1, Cit2	11658	84194	47	33	477	52946	2539
18	AER031C	Tdh3	10779	33108	43	20	331	35479	2601
19	AER184W	Pet9, Aac3	9928	41193	33	20	305	33110	1886
20	AFR279C	Hxk2, Hxk1	9817	75713	53	29	488	53943	3267

Histone occurrence and acetylation in *A. gossypii* spores

Spores carry one nucleus with a genome of $9.1x10^6$ bp. We anticipated that the majority of the genomic DNA in spores is organized in nucleosomes, with the DNA wound around histone octamers (Richmond and Davey, 2003; Wilmen and Hegemann, Chromosoma 1996). Because of the established composition of nucleosomes, histones are expected to be present in similar amounts also in spores. Like in *S. cerevisiae*, each of the four core histones H2A, H2B, H3 and H4 is encoded by two genes in *A. gossypii*, whereas the evolutionary conserved H2A-variant Htz1 is encoded by a single gene in both fungi.

The relative abundances of the four *A. gossypii* core histones as determined by the All-method or Top3-method (Table2-2 Appendix 1) showed unexpected differences, e.g. the abundance of H3 was only 6% of the H4 abundance. Therefore we tried a third approach to determine the relative abundance of histones. We first documented for each histone all peptides identified by the tandem MS/MS analyses, including partially cleaved peptides, acetylated peptides, differently charged peptides, and peptides found in different fractions of the OffGel separation. These automated outputs also list the measured abundances of the detected peptides. Next we marked in the individual histone sequences the detected peptides and lysines modified by acetylation. Than we compiled in a subset of tables color-coded highly abundant peptides in order to mark for each histone the most abundant peptide sequence. These tables also list longer peptides originating from partial tryptic digestions or acetylated lysines when they contain the sequence of a selected peptide. During this search we occasionally included peptides classified as FALSE when the software had classified the same peptide in a different fraction as TRUE. Finally, we compare in a summary table the relative histone abundances found by the two automated methods and the most abundant peptide approach. We also compare the N-terminal lysine acetylation patterns of *A. gossypii* and *S. cerevisiae* histones.

Table2-4: *Saccharomyces* Genome Database (SGD) description of the *S. cerevisiae* homologs for the top 20 *A. gossypii* spore proteins using the All-Method.

no.	accession	common name	name description	SGD description of the firth *S. cerevisiae* homolog
1	AAL128C	Hsp26	Heat Shock Protein	Small heat shock protein (sHSP) with chaperone activity; forms hollow, sphere-shaped oligomers that suppress unfolded proteins aggregation; oligomer activation requires heat-induced conformational change; also has mRNA binding activity
2	ADL370C	Tef1, Tef2	Translation Elongation Factor	Translational elongation factor EF-1 alpha; also encoded by TEF2; functions in the binding reaction of aminoacyl-tRNA (AA-tRNA) to ribosomes; may also have a role in tRNA re-export from the nucleus
3	AGL321W	Sod1	SuperOxide Dismutase	Cytosolic copper-zinc superoxide dismutase; some mutations are analogous to those that cause ALS (amyotrophic lateral sclerosis) in humans
4	ADL154C	NOHBY412	No homolog in Saccharomyces cerevisiae	
5	AFR694W	YLR179C, Tfs1	cdc25 (Twenty-Five) Suppressor	Protein that interacts with and inhibits carboxypeptidase Y and Ira2p; phosphatidylethanolamine-binding protein (PEBP) family member; targets to vacuolar membranes during stationary phase; acetylated by NatB N-terminal acetyltransferase
6	AER312W	Tsa1, Tsa2	Thiol-Specific Antioxidant	Thioredoxin peroxidase, acts as both a ribosome-associated and free cytoplasmic antioxidant; self-associates to form a high-molecular weight chaperone complex under oxidative stress; deletion results in mutator phenotype
7	ABL156C	Fmp45	Found in Mitochondrial Proteome	Integral membrane protein localized to mitochondria (untagged protein); required for sporulation and maintaining sphingolipid content; has sequence similarity to SUR7 and YNL194C
8	ABR068C	Fba1	Fructose 1,6-bisphosphate aldolase	Fructose 1,6-bisphosphate aldolase, required for glycolysis and gluconeogenesis; catalyzes conversion of fructose 1,6 bisphosphate to glyceraldehyde-3-P and dihydroxyacetone-P; locates to mitochondrial outer surface upon oxidative stress
9	ADL032W	Aco1	ACOnitase	Aconitase, required for the tricarboxylic acid (TCA) cycle and also independently required for mitochondrial genome maintenance; phosphorylated; component of the mitochondrial nucleoid; mutation leads to glutamate auxotrophy
10	ABR028C	Cwp1	Cell Wall Protein	Cell wall mannoprotein that localizes specifically to birth scars of daughter cells, linked to a beta-1,3- and beta-1,6-glucan heteropolymer through a phosphodiester bond; required for propionic acid resistance
11	ABR222W	Act1	ACTin	Actin, structural protein involved in cell polarization, endocytosis, and other cytoskeletal functions
12	AFR020W	Ach1	Acetyl CoA Hydrolase	Protein with CoA transferase activity, particularly for CoASH transfer from succinyl-CoA to acetate; has minor acetyl-CoA-hydrolase activity; phosphorylated; required for acetate utilization and for diploid pseudohyphal growth
13	AGL177C	Cpr1	Cyclosporin A-sensitive Proline Rotamase	Cytoplasmic peptidyl-prolyl cis-trans isomerase (cyclophilin), catalyzes the cis-trans isomerization of peptide bonds N-terminal to proline residues; binds the drug cyclosporin A
14	AER428W	Om45	Outer Membrane	Protein of unknown function, major constituent of the mitochondrial outer membrane; located on the outer (cytosolic) face of the outer membrane
15	ACL198W	Ilv5	soLeucine-plus-Valine requiring	Bifunctional acetohydroxyacid reductoisomerase and mtDNA binding protein; involved in branched-chain amino acid biosynthesis and maintenance of wild-type mitochondrial DNA; found in mitochondrial nucleoids
16	AER401W	Ypr1, Gcy1	Yeast Putative Reductase	NADPH-dependent aldo-keto reductase, reduces multiple substrates including 2-methylbutyraldehyde and D,L-glyceraldehyde, expression is induced by osmotic and oxidative stress; functionally redundant with other aldo-keto reductases
17	AAR004C	Cit1, Cit2	CITrate synthase	Citrate synthase, catalyzes the condensation of acetyl coenzyme A and oxaloacetate to form citrate; the rate-limiting enzyme of the TCA cycle; nuclear encoded mitochondrial protein
18	AER031C	Tdh3	Triose-phosphate DeHydrogenase	Glyceraldehyde-3-phosphate dehydrogenase, isozyme 3, involved in glycolysis and gluconeogenesis; tetramer that catalyzes the reaction of glyceraldehyde-3-phosphate to 1,3 bis-phosphoglycerate; detected in the cytoplasm and cell wall
19	AER184W	Pet9, Aac3	PETite colonies	Major ADP/ATP carrier of the mitochondrial inner membrane, exchanges cytosolic ADP for mitochondrially synthesized ATP; also imports heme and ATP; phosphorylated; required for viability in many lab strains that carry a sal1 mutation
20	AFR279C	Hxk2, Hxk1	Hexokinase isoenzyme 2	Hexokinase isoenzyme 2 that catalyzes phosphorylation of glucose in the cytosol; predominant hexokinase during growth on glucose; functions in the nucleus to repress expression of HXK1 and GLK1 and to induce expression of its own gene

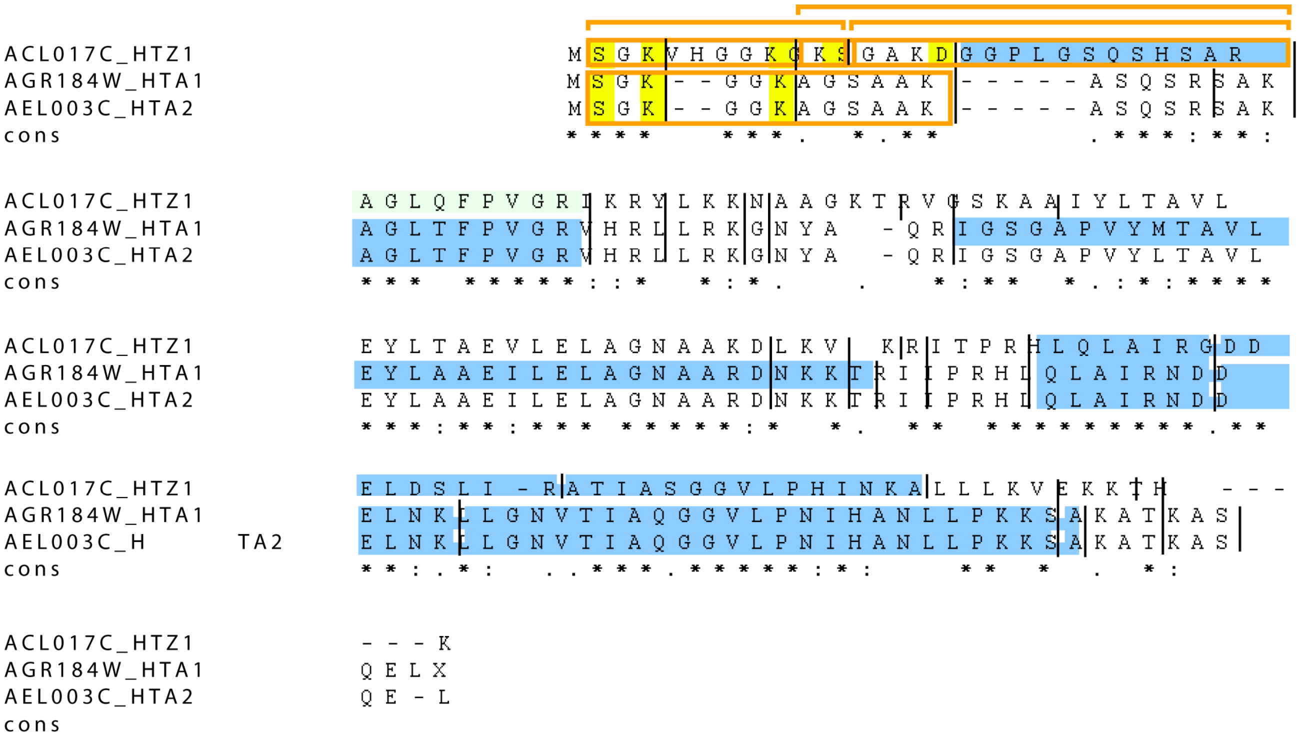
Histones Hta1, Hta2 and Htz1

Histone H2A is a core histone and forms heterodimers with histone H2B (Schafer et al., 2005). *A. gossypii* encodes two genes for histone H2A. In the following, we use the SGD nomenclature Hta1 and Hta2 for the two *A. gossypii* H2A histones. Both copies are highly similar (Figure2-10) also with the two *S. cerevisiae* homologs (data not shown). Peptides detected via tandem MS (MS/MS) are listed in Table2-5 and are marked in blue in Figure2-10. As explained in the technical Chapter peptides with a length of more than six to seven amino acids (molecular weight beyond 700 Dalton) can be detected by LTQ Orbitrap. Shorter tryptic peptides from the N-terminus of Hta1 and Hta2 became detectable when the lysines at position 3 and 6 (mature protein) were acetylated which blocked cleavage by trypsin yielding longer peptides as indicated by red brackets in Figure2-10. Despite the high sequence similarity between Hta1 and Hta2, one unique peptide from the center should be detectable for both histones. Only one of these specific peptides was found. Therefore, Hta1 and Hta2 could not be separately quantified. The most abundant peptide (LLGNVTIAQGGVLPNIHANLLPK) of Hta1 and Hta2 is marked in red (Table2-7) with a relative abundance of 10.9x10^{6}. The sequence of this peptide contains a positively charged histidine and no negatively charged amino acid side chains, which favors ionization and therefore more frequent detection compared to other tryptic peptides. Additionally, we found this peptide sequence in three partially cleaved peptides. The abundances of all these peptides were added to the peptide sequence "LLGNVTIAQGGVLPNIHANLLPK" which increases its relative abundance to 12.5 million (bottom of Table2-7). The peptide HLQLAIR has an even higher abundance of 22.7 million but it could also originate from Htz1 (see Figure2-10). Assuming that the sum of Hta1 and Hta2 abundances is significantly higher than the Htz1 abundance one can estimate that 80% to 90% of HLQLAIR originate from Hta1 and Hta2 as discussed below.

Table2-5: MASCOT output of the detected peptides and measured abundances of the two *A. gossypii* H2A histones. The table lists for each peptide the sequence, the modification, the normalized abundance, the score, and the fraction of the OffGel-electrophoresis in which the peptide was detected. The validation "TRUE" or "FALSE" indicates whether the peptide was included or excluded for the automated calculation of the protein abundance.

common name	Sequence	Modifications	0h Normalized abundance
Hta1	IGSGAPVYMTAVLEYLAAEILELAGNAARDNKK	[9] Oxidation (M)	6531
Hta1 / Hta2	AGLTFPVGR		222701
Hta1 / Hta2	HLQLAIRNDDELNK		148909
Hta1 / Hta2	HLQLAIRNDDELNKLLGNVTIAQGGVLPNIHANLLPK		29306
Hta1 / Hta2	LLGNVTIAQGGVLPNIHANLLPK		10884117
Hta1 / Hta2	NDDELNK		1852
Hta1 / Hta2	NDDELNKLLGNVTIAQGGVLPNIHANLLPK		1467171
Hta1 / Hta2	NDDELNKLLGNVTIAQGGVLPNIHANLLPKK		76802
Hta1 / Hta2	SGKGGKAGSAAK	[N-term] Acetyl (Protein N-term)|[3] Acetyl (K)|[6] Acetyl (K)	2787883
Htz1	ATIASGGVLPHINK		1330
Htz1	DGGPLGSQSHSAR		2676668
Htz1	GKSGAKDGGPLGSQSHSAR	[2] Acetyl (K)|[6] Acetyl (K)	396643
Htz1	HLQLAIRGDDELDSLIR		87955
Htz1	HLQLAIRGDDELDSLIRATIASGGVLPHINK		821
Htz1	SGAKDGGPLGSQSHSAR	[4] Acetyl (K)	176814
Htz1	SGKVHGGKGK	[N-term] Acetyl (Protein N-term)|[3] Acetyl (K)|[8] Acetyl (K)	192564
Hta1 / Hta2 / Htz1	HLQLAIR		22456326
Hta1 / Hta2	LLGNVTIAQGGVLPNIHANLLPK		**12457397**
Htz1	DGGPLGSQSHSAR		**3073311**
Hta1 / Hta2 / Htz1	HLQLAIR		**22723316**

Histone Htz1 is a variant of Hta1 and Hta2 and can replace these histones in a subset of nucleosomes (Jackson et al., 1996). Histones Hta1 and Hta2 in *A. gossypii* show sequence similarities to Htz1 (Figure2-10). Table2-6 lists the Htz1 peptide output of the MASCOT search with abundances and scores. The peptide coverage of Htz1, marked in blue in Figure2-10, is higher compared to Hta1 and Hta2. One peptide marked in light green has a MASCOT score of less than 15 and was therefore not included for the quantifications. Four acetylation sites were detected at the N-terminus of Htz1 (AcK3, AcK8, AcK10 and AcK14). The DGGPLGSQSHSAR peptide, marked in blue in Table2-7, is the most abundant Htz1 peptide most likely because of a high ionization potential (presence of one histidine and lack of negatively charged amino acids). Together with the acetylated peptide variants of GKSGAKDGGPLGSQSHSAR, we calculated a total abundance of 3.22×10^6 (Table2-7).



```
ACL017C_HTZ1                          M S G K V H G G K G K S G A K D G G P L G S Q S H S A R
AGR184W_HTA1                          M S G K - - G G K A G S A A K - - - - - A S Q S R S A K
AEL003C_HTA2                          M S G K - - G G K A G S A A K - - - - - A S Q S R S A K
cons                                  * * * *     * * * .   * . * *           . * * * : * * :

ACL017C_HTZ1        A G L Q F P V G R I K R Y L K K N A A G K T R V G S K A A I Y L T A V L
AGR184W_HTA1        A G L T F P V G R V H R L L R K G N Y A   - Q R I G S G A P V Y M T A V L
AEL003C_HTA2        A G L T F P V G R V H R L L R K G N Y A   - Q R I G S G A P V Y L T A V L
cons                * * *   * * * * * : : *     * : * .     .     * : * *   * . : * : * * * *

ACL017C_HTZ1        E Y L T A E V L E L A G N A A K D L K V   K R I T P R H L Q L A I R G D D
AGR184W_HTA1        E Y L A A E I L E L A G N A A R D N K K T R I I P R H L Q L A I R N D D
AEL003C_HTA2        E Y L A A E I L E L A G N A A R D N K K T R I I P R H L Q L A I R N D D
cons                * * * : * * : * * *   * * * * * : *     * .   * *   * * * * * * * * * . * *

ACL017C_HTZ1        E L D S L I - R A T I A S G G V L P H I N K A L L L K V E K K T H     - - -
AGR184W_HTA1        E L N K L L G N V T I A Q G G V L P N I H A N L L P K K S A K A T K A S
AEL003C_H      TA2  E L N K L L G N V T I A Q G G V L P N I H A N L L P K K S A K A T K A S
cons                * * : . * :     . . * * * . * * * * * : * :         * *     *     .     * :

ACL017C_HTZ1        - - - K
AGR184W_HTA1        Q E L X
AEL003C_HTA2        Q E - L
cons
```

Figure2-10: Sequence alignment of Htz1 with the two H2A histones and experimentally identified peptides. Peptides detected by tandem mass spectrometry are marked in blue. Lysines highlighted in yellow can be acetylated resulting in detected peptides marked by orange brackets. Peptides with a score less than 15 were not used for protein quantification and marked in light green.

There are two numbers for the estimation of the Hta1/Hta2 abundance: 12.5 million based on the most abundant Hta1/Hta2 peptide and 22.7 million for the most abundant peptide found for the sum of Hta1, Hta2, and Htz1. There is a fairly reliable number for the Htz1 abundance of 3.2 million based on the most abundant peptide with a high ionization potential and the high coverage of Htz1 by detected peptides. Therefore the abundance of Hta1/Hta2 can be estimated as 19. 5 million.

Table2-6: MASCOT output of the detected peptides and measured abundances of the *A. gossypii* H2A-like histone Htz1. The table list for each peptide the sequence, the modification, the normalized abundance, the score, and the fraction of the OffGel-electrophoresis in which the peptide was detected. The validation "TRUE" or "FALSE" indicates whether the peptide was included or excluded for the automated calculation of the protein abundance.

Accession	common name	Sequence	Modifications	0h Normalized abundance	Use in quantitation	Score	Fraction
ACL017C	Htz1	GKSGAKDGGPLGSQSHSAR	[2] Acetyl (K)|[6] Acetyl (K)	48030.47342	TRUE	23.89	8
ACL017C	Htz1	GKSGAKDGGPLGSQSHSAR	[2] Acetyl (K)|[6] Acetyl (K)	348612.7854	TRUE	41.37	9
ACL017C	Htz1	SGAKDGGPLGSQSHSAR	[4] Acetyl (K)	176814.2788	TRUE	38.5	8
ACL017C	Htz1	SGKVHGGKGK	[N-term] Acetyl (Protein N-term)|[3] Acetyl (K)|[8] Acetyl (K)	192564.1557	TRUE	42.64	10
ACL017C	Htz1	HLQLAIR		10863057.76	FALSE	37.44	12
ACL017C	Htz1	HLQLAIR		2152.123929	FALSE	35.93	12
ACL017C	Htz1	ATIASGGVLPHINK		1329.571362	TRUE	98.54	12
ACL017C	Htz1	HLQLAIR		508.1979907	FALSE	21.83	10
ACL017C	Htz1	DGGPLGSQSHSAR		1877390.403	FALSE	63.33	8
ACL017C	Htz1	DGGPLGSQSHSAR		799277.6334	TRUE	18.83	8
ACL017C	Htz1	HLQLAIRGDDELDSLIRATIASGGVLPHINK		820.591816	TRUE	25.34	5
ACL017C	Htz1	HLQLAIRGDDELDSLIR		83626.02079	TRUE	37.92	3
ACL017C	Htz1	HLQLAIRGDDELDSLIR		4328.593404	TRUE	20.15	3

Table2-7: Compilation of all highly abundant peptides of histones Hta1, Hta2, and Htz1. Color codes facilitate to visualize the same sequence in completely or partially cleaved peptides. At the bottom of the table the three most abundant peptides are listed.

common name	Sequence	Modifications	0h Normalized abundance
Hta1	IGSGAPVYMTAVLEYLAAEILELAGNAARDNKK	[9] Oxidation (M)	6531
Hta1 / Hta2	AGLTFPVGR		222701
Hta1 / Hta2	**HLQLAIR**NDDELNK		148909
Hta1 / Hta2	**HLQLAIR**NDDELNK**LLGNVTIAQGGVLPNIHANLLPK**		29306
Hta1 / Hta2	**LLGNVTIAQGGVLPNIHANLLPK**		10884117
Hta1 / Hta2	NDDELNK		1852
Hta1 / Hta2	NDDELNK**LLGNVTIAQGGVLPNIHANLLPK**		1467171
Hta1 / Hta2	NDDELNK**LLGNVTIAQGGVLPNIHANLLPK**K		76802
Hta1 / Hta2	SGKGGKAGSAAK	[N-term] Acetyl (Protein N-term)|[3] Acetyl (K)|[6] Acetyl (K)	2787883
Htz1	ATIASGGVLPHINK		1330
Htz1	**DGGPLGSQSHSAR**		2676668
Htz1	GKSGAK**DGGPLGSQSHSAR**	[2] Acetyl (K)|[6] Acetyl (K)	396643
Htz1	**HLQLAIR**GDDELDSLIR		87955
Htz1	**HLQLAIR**GDDELDSLIRATIASGGVLPHINK		821
Htz1	SGAKDGGPLGSQSHSAR	[4] Acetyl (K)	176814
Htz1	SGKVHGGKGK	[N-term] Acetyl (Protein N-term)|[3] Acetyl (K)|[8] Acetyl (K)	192564
Hta1 / Hta2 / Htz1	**HLQLAIR**		22456326
Hta1 / Hta2	**LLGNVTIAQGGVLPNIHANLLPK**		**12457397**
Htz1	**DGGPLGSQSHSAR**		**3073311**
Hta1 / Hta2 / Htz1	**HLQLAIR**		**22723316**

Table2-8: MASCOT output of the detected peptides and measured abundances of the two *A. gossypii* H2B histones. The table list for each peptide the sequence, the modification, the normalized abundance, the score, and the fraction of the OffGel-electrophoresis in which the peptide was detected. The validation "TRUE" or "FALSE" indicates whether the peptide was included or excluded for the automated calculation of the protein abundance.

Accession	common name	Sequence	Modifications	0h Normalized abundance	Use in quantitation	Score	Fraction
AEL002W	Htb2	SMSILNSFVNDIFERIASEASK	[2] Oxidation (M)	5642.894796	TRUE	42.71	3
AEL002W	Htb2	SMSILNSFVNDIFER	[2] Oxidation (M)	184002.4326	TRUE	89.63	3
AEL002W	Htb2	SMSILNSFVNDIFER	[2] Oxidation (M)	10294047.47	TRUE	105.36	2
AEL002W	Htb2	SMSILNSFVNDIFER	[2] Oxidation (M)	172612.6474	TRUE	65.01	2
AEL002W	Htb2	AEKKPASK	[4] Acetyl (K)	7148716.595	FALSE	23.58	12
AEL002W	Htb2	APAAKKTTASTDASK	[5] Acetyl (K)|[6] Acetyl (K)	0	TRUE	61.87	10
AEL002W	Htb2	APAAKKTTASTDASK	[5] Acetyl (K)|[6] Acetyl (K)	0	TRUE	74.37	8
AEL002W	Htb2	APAAKKTTASTDASK	[5] Acetyl (K)|[6] Acetyl (K)	0	TRUE	83.27	7
AEL002W	Htb2	APAAKKTTASTDASK	[5] Acetyl (K)|[6] Acetyl (K)	6199.132514	TRUE	78.09	6
AEL002W	Htb2	LILPGELAK	[9] Acetyl (K)	71765.4232	FALSE	34.68	6
AEL002W	Htb2	VLKQTHPDTGISQK		190.078968	TRUE	47.35	12
AEL002W	Htb2	VLKQTHPDTGISQK		636.0996222	TRUE	66.17	12
AEL002W	Htb2	VLKQTHPDTGISQK		14.14628425	FALSE	34.47	12
AEL002W	Htb2	KETYSSYIYK		226474.4617	TRUE	34.64	12
AEL002W	Htb2	QTHPDTGISQK		16.05298475	TRUE	25.73	12
AEL002W	Htb2	KETYSSYIYK		9863758.138	TRUE	39.67	11
AEL002W	Htb2	KETYSSYIYK		1471811.484	TRUE	29.08	11
AEL002W	Htb2	VLKQTHPDTGISQK		10927.06647	TRUE	26.63	11
AEL002W	Htb2	ETYSSYIYKVLK		0	TRUE	31.69	11
AEL002W	Htb2	ETYSSYIYKVLK		0	TRUE	61.06	11
AEL002W	Htb2	QTHPDTGISQK		244.3463052	TRUE	21.41	11
AEL002W	Htb2	QTHPDTGISQK		26.79412039	TRUE	24.76	11
AEL002W	Htb2	ETYSSYIYK		17845.09298	FALSE	22.85	10
AEL002W	Htb2	QTHPDTGISQK		26211.73582	FALSE	28.82	10
AEL002W	Htb2	KETYSSYIYK		626561.4379	FALSE	39.22	10
AEL002W	Htb2	QTHPDTGISQK		805298.3757	TRUE	48.04	8
AEL002W	Htb2	QTHPDTGISQK		1000305.834	TRUE	27.24	8
AEL002W	Htb2	QTHPDTGISQK		1661.563675	TRUE	51.05	8
AEL002W	Htb2	QTHPDTGISQK		3590.636885	TRUE	27.04	8
AEL002W	Htb2	LILPGELAK		616107.4812	TRUE	24.32	8
AEL002W	Htb2	ETYSSYIYK		0	TRUE	19.36	8
AEL002W	Htb2	LILPGELAK		2103926.448	TRUE	22.81	7
AEL002W	Htb2	QTHPDTGISQK		13985.05173	TRUE	53.38	7
AEL002W	Htb2	QTHPDTGISQK		4806.73183	TRUE	44.02	7
AEL002W	Htb2	EIQTAVR		216866.2514	TRUE	30.76	7
AEL002W	Htb2	LILPGELAK		22757822.49	FALSE	21.59	6
AEL002W	Htb2	EIQTAVR		525516.7177	FALSE	28.41	6
AEL002W	Htb2	ETYSSYIYK		0	FALSE	29.88	6
AEL002W	Htb2	ETYSSYIYK		596876.776	FALSE	26.29	6
AEL002W	Htb2	EIQTAVRLILPGELAK		17840.47141	FALSE	19.33	6
AEL002W	Htb2	EIQTAVR		127.5409392	FALSE	29.41	4
AEL002W	Htb2	ETYSSYIYK		408021.9178	TRUE	18.36	4
AEL002W	Htb2	QTHPDTGISQK		4902.536315	TRUE	27.2	9
AEL002W	Htb2	QTHPDTGISQK		48746.82321	TRUE	16.94	9
AEL002W	Htb2	LILPGELAK		10012.96444	TRUE	22.3	9
AEL002W	Htb2	ETYSSYIYK		25604.53025	TRUE	19.45	9
AEL002W	Htb2	SMSILNSFVNDIFER		141968.1391	FALSE	102.38	2
AEL002W	Htb2	SMSILNSFVNDIFER		5661557.571	TRUE	55.69	2
AEL002W	Htb2	SMSILNSFVNDIFER		819.6030402	TRUE	52.46	2
AEL002W	Htb2	SMSILNSFVNDIFER		120366.571	TRUE	49.9	2
AGR183C	Htb1	ASKAPASKAPAEK	[3] Acetyl (K)|[8] Acetyl (K)	348.0092243	TRUE	46.88	6
AGR183C	Htb1	ASKAPASKAPAEK	[3] Acetyl (K)|[8] Acetyl (K)	188671.3645	TRUE	39.4	6
AGR183C	Htb1	APASKAPAEK	[5] Acetyl (K)	343.3552317	TRUE	51.9	6
AGR183C	Htb1	LILPGELAK	[9] Acetyl (K)	71765.4232	FALSE	34.68	6
AGR183C	Htb1	SSKASKAPASK	[N-term] Acetyl (Protein N-term)|[3] Acetyl (K)|[6] Acetyl (K)	2638.56879	TRUE	34.97	10
AGR183C	Htb1	ETYSSYIYK		17845.09298	FALSE	22.85	10
AGR183C	Htb1	QTHPDTGISQK		26211.73582	FALSE	28.82	10
AGR183C	Htb1	KETYSSYIYK		626561.4379	FALSE	39.22	10
AGR183C	Htb1	LILPGELAK		22757822.49	FALSE	21.59	6
AGR183C	Htb1	EIQTAVR		525516.7177	FALSE	28.41	6
AGR183C	Htb1	ETYSSYIYK		0	FALSE	29.88	6
AGR183C	Htb1	ETYSSYIYK		596876.776	FALSE	26.29	6
AGR183C	Htb1	EIQTAVRLILPGELAK		17840.47141	FALSE	19.33	6
AGR183C	Htb1	LILPGELAK		2696144.94	TRUE	31.93	5
AGR183C	Htb1	ETYSSYIYK		1393180.032	TRUE	29.43	5
AGR183C	Htb1	EIQTAVR		263.4928265	TRUE	28.67	5
AGR183C	Htb1	EIQTAVR		621335.712	TRUE	27.66	5

Histones Htb1 and Htb2

It is well established that histones H2A and H2B interact as a heterodimer and therefore should be found at similar abundances. Like histone H2A discussed in the previous section histone 2B is encoded by two gene copies in *A. gossypii*. The names Htb1 and Htb2 are based on the gene names according to the rules of SGD. All detected Htb1 and Htb2 peptides with MASCOT score > 15 are listed in Table2-8. *A. gossypii* Htb1 and Htb2 sequences are very similar (Figure2-11) and also very similar to their *S. cerevisiae* homologs (not shown) with a conservation T-Coffee score of 98 (max score = 100). Most detectable peptides were identified and are marked in blue in Figure2-11. Htb1 and Htb2 carry three conserved N-terminal lysines, three of which are acetylated in Htb1 but only one in Htb2 (Table2-8). The longer tryptic peptides generated due to the lysine acetylations are indicated by red brackets in Figure2-11. Htb2 carries two lysines at positions 17 and 18, both of which can be acetylated (Table2-8, Figure2-11). These two lysins are absent in Htb1. Interestingly, we could identify an amino-terminal peptide of Htb2 with the start methionine (Table2-8). Due to the low MASCOT peptide score (< 15), we did not use this peptide for further quantifications.

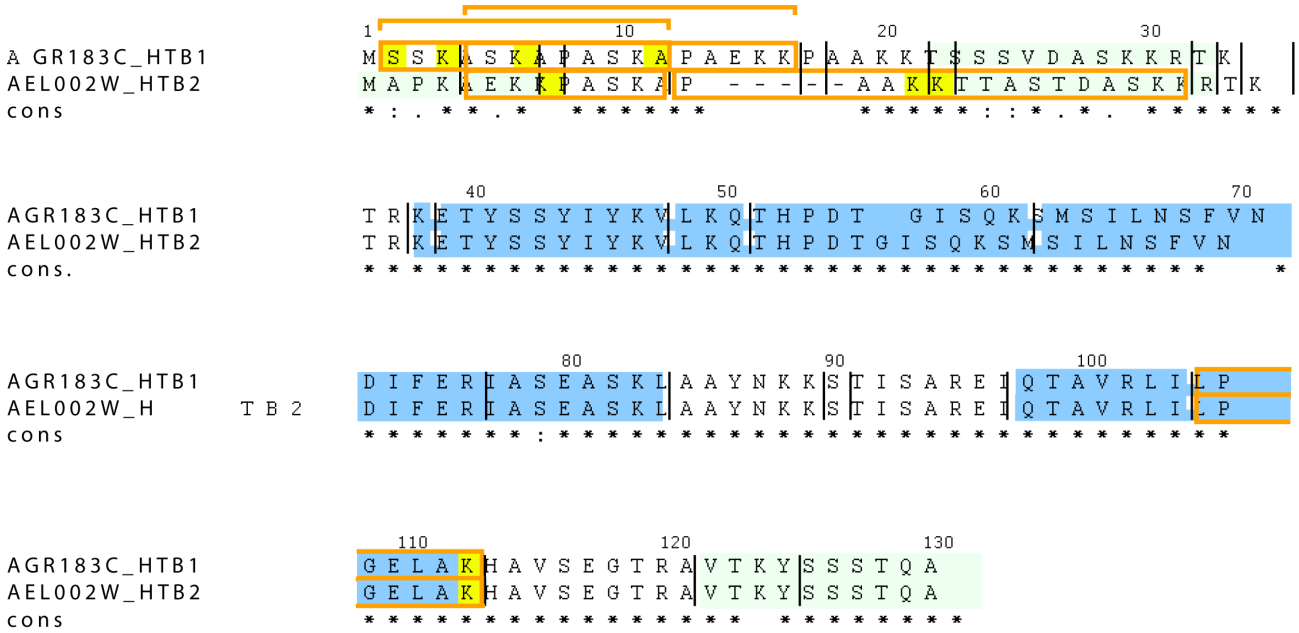


Figure2-11: Sequence alignment of both H2B histones and experimentally identified peptides. Detected peptides are marked in blue. Lysines highlighted in yellow can be acetylated resulting in detected peptides marked by orange brackets. Peptides with a score less than 15 were not used for protein quantification and marked in light green.

The high sequence similarity of Htb1 and Htb2 prevents an individual determination of the abundance for each protein. Like for Hta1 and Hta2 the highest abundant peptides for Htb1 and Htb2 were compiled in a table highlighting the three top peptides (Table2-9). LILPGELAK is the highest abundant peptide, marked in red, which was also detected in an acetylated variant. In total, the peptide LILPGELAK has a relative abundance of 51.1x10^6.

In conclusion, a reliable estimation of the relative abundance of Htb1 and Htb2 is based on a high coverage of detected peptides and several highly abundant peptides. Based on the highest abundant peptide Htb1 and Htb2 together have an abundance of 51 million.

Table2-9: Compilation of all highly abundant peptide of histone Htb1 and Htb2. Color codes facilitate to visualize the same sequence in completely or partially cleaved peptides. At the bottom of the table the three most abundant peptides are listed.

histone	common name	Sequence	Modifications	0h Normalized abundance
H2B	Htb1 / Htb2	EIQTAVR		1889626
H2B	Htb1 / Htb2	EIQTAVRLILPGELAK		35681
H2B	Htb1 / Htb2	**ETYSSYIYK**		3056250
H2B	Htb1 / Htb2	**ETYSSYIYK**VLK		0
H2B	Htb1 / Htb2	K**ETYSSYIYK**		12815167
H2B	Htb1 / Htb2	**LILPGELAK**	[9] Acetyl (K)	143531
H2B	Htb1 / Htb2	**LILPGELAK**		50941837
H2B	Htb1 / Htb2	QTHPDTGISQK		1936008
H2B	Htb1 / Htb2	**SMSILNSFVNDIFER**		16575374
H2B	Htb1 / Htb2	**SMSILNSFVNDIFER**IASEASK	[2] Oxidation (M)	5643
H2B	Htb1 / Htb2	VLKQTHPDTGISQK		11767
H2B	Htb2	APASKAPAEK	[5] Acetyl (K)	343
H2B	Htb2	ASKAPASKAPAEK	[3] Acetyl (K)|[8] Acetyl (K)	189019
H2B	Htb1	AEKKPASK	[4] Acetyl (K)	7148717
H2B	Htb1	APAAKKTTASTDASK	[5] Acetyl (K)|[6] Acetyl (K)	6199
H2B	Htb1	SSKASKAPASK	[N-term] Acetyl (Protein N-term)|[3] Acetyl (K)|[6] Acetyl (K)	2639
		LILPGELAK		**51085368**
		SMSILNSFVNDIFER		**16581017**
		ETYSSYIYK		**15871417**

Histone 3 and Histone 4

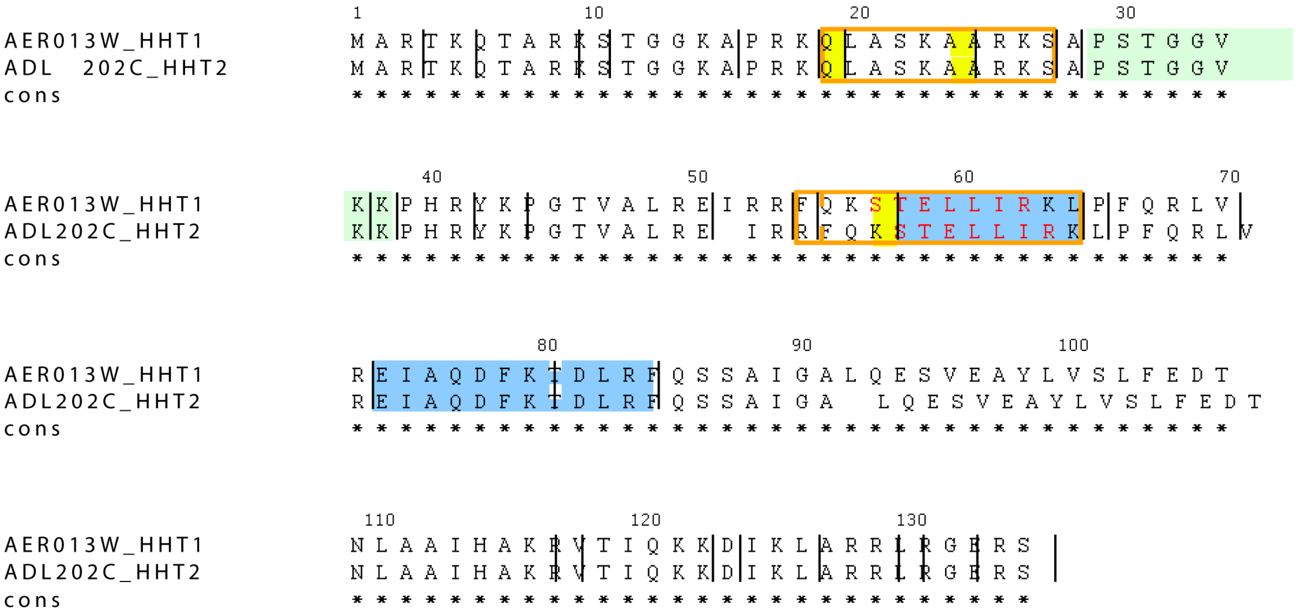
Histones H3 and H4 interact as a heterodimer, and two of these heterodimers form together with two heterodimers of histone H2A and H2B the histone octamer of nucleosomes (Schafer et al., 2005). Both genes for histone H3 (HHT1 and HHT2) encode identical proteins (Figure2-12). Interestingly, the sequence is also identical to both *S. cerevisiae* homologs. The two H3 histones are called Hht1 and Hht2 according to the SGD rules. Histone H4 is also encoded by two genes (HHF1 and HHF2). The two proteins are called Hhf1 and Hhf2 and have identical sequences (Figure2-13) which are also highly conserved in the *S. cerevisiae* homologs (data not shown).

The automatic output of the identified peptides using MASCOT search for Hht1 and Hht2 is listed in Table2-10. It reveals three acetylation sites at lysine residues 19, 24, and 57 but only a very low overall sequence coverage by two peptides, marked in blue, because digestion with trypsin generates either too long or too short peptides. The most abundant peptide is STELLIR, marked in red in Tabel2-11. Together with two acetylated variants the total abundance of this peptide sequence adds up to 12.1x10^6.

Table2-10: MASCOT output of the detected peptides and measured abundances of the two *A. gossypii* H3 histones. The table list for each peptide the sequence, the modification, the normalized abundance, the score, and the fraction of the OffGel-electrophoresis in which the peptide was detected. The validation "TRUE" or "FALSE" indicates whether the peptide was included or excluded for the automated calculation of the protein abundance.

Accession	common name	Sequence	Modifications	0h Normalized abundance	Use in quantitation	Score	Fraction
ADL202C	Hht2	KQLASKAAR	[1] Acetyl (K)|[6] Acetyl (K)	238735.0058	TRUE	42.04	12
ADL202C	Hht2	KQLASKAAR	[1] Acetyl (K)|[6] Acetyl (K)	18905.85295	TRUE	18.7	10
ADL202C	Hht2	FQKSTELLIR	[3] Acetyl (K)	2953502.447	TRUE	43.2	5
ADL202C	Hht2	RFQKSTELLIR	[4] Acetyl (K)	23851.93912	TRUE	29.34	12
ADL202C	Hht2	RFQKSTELLIR	[4] Acetyl (K)	5657.927131	TRUE	52.11	11
ADL202C	Hht2	RFQKSTELLIR	[4] Acetyl (K)	738.8100378	FALSE	34.36	10
ADL202C	Hht2	RFQKSTELLIR	[4] Acetyl (K)	2348.277102	TRUE	23.88	10
ADL202C	Hht2	EIAQDFKTDLR		7412.025889	TRUE	15.95	1
ADL202C	Hht2	EIAQDFKTDLR		494764.8167	TRUE	37.01	1
ADL202C	Hht2	EIAQDFK		1121412.296	TRUE	18.24	1
ADL202C	Hht2	EIAQDFK		753.3373829	TRUE	25.25	1
ADL202C	Hht2	STELLIR		14014.94873	FALSE	35.07	10
ADL202C	Hht2	STELLIR		9049356.127	FALSE	50.25	5
ADL202C	Hht2	STELLIR		51926.81543	FALSE	53.34	5
ADL202C	Hht2	STELLIR		271.717404	FALSE	50.65	4

The analysis of Hhf1 and Hhf2 showed a very high sequence coverage by detected peptides (Table2-12 and Figure2-13). One acetylation site was found at lysine17. The highest abundant peptide is DNIQGITK, marked in red in Table2-13 that was also present in a partially cleaved peptide. The sum of both abundances is 46,7x10^6. It is worth mentioning that the peptide ISGLIYEDVR, marked in yellow in Table2-13, was found with a similar abundance, and that the abundance of two other peptides is higher than the most abundant Hta1/Hta2 and Hht1/Hht2 peptides.



```
                  1                       10                         20                         30
AER013W_HHT1      M A R T K Q T A R K S T G G K A P R K Q L A S K A A R K S A P S T G G V
ADL  202C_HHT2    M A R T K Q T A R K S T G G K A P R K Q L A S K A A R K S A P S T G G V
cons              * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

                          40                         50                         60                         70
AER013W_HHT1      K K P H R Y K P G T V A L R E I R R F Q K S T E L L I R K L P F Q R L V
ADL202C_HHT2      K K P H R Y K P G T V A L R E   I R R F Q K S T E L L I R K L P F Q R L V
cons              * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

                                      80                         90                         100
AER013W_HHT1      R E I A Q D F K T D L R F Q S S A I G A L Q E S V E A Y L V S L F E D T
ADL202C_HHT2      R E I A Q D F K T D L R F Q S S A I G A   L Q E S V E A Y L V S L F E D T
cons              * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

                    110                        120                        130
AER013W_HHT1      N L A A I H A K R V T I Q K K D I K L A R R L R G E R S
ADL202C_HHT2      N L A A I H A K R V T I Q K K D I K L A R R L R G E R S
cons              * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

Figure2-12: Sequence alignment of the two H3 histones H3 and experimentally identified peptides. Detected peptides are marked in blue. Lysines highlighted in yellow can be acetylated resulting in detected peptides marked by orange brackets. Peptides with a score less than 15 were not used for protein quantification and marked in light green. The AQUA peptide used in subsequent experiments is marked in red letters (see chapter II-C).

In conclusion, the very low coverage of Hht1/Hht2 by detected peptides very likely leads to an underestimation of their relative abundance based on one peptide with an abundance of 12 million. On the other hand the estimation of the relative abundance of 47 million for Hhf1/Hhf2 is most likely reliable because it is based on a high coverage of detected peptides and on two peptides with very high abundances.

If one assumes that the abundance of a protein is as high or higher than the most abundant peptide we can define the minimal relative abundances of both histones in spores: 12 million for histone H3 (Hht1 and Hht2) and 47 million for histone H4 (Hhf1 and Hhf2). This results in a stoichiometric ratio of 1:4 and not the expected ratio of 1:1. Possible reasons for this deviation will be discussed below.

Table2-11: Compilation of all highly abundant peptides of histones Hht1 and Hht2. Color codes facilitate to visualize the same sequence in completely or partially cleaved peptides. At the bottom of the table the most abundant peptide is listed.

histone	common name	Sequence	Modifications	0h Normalized abundance
H3	Hht1 / Hht2	KQLASKAAR	[1] Acetyl (K)|[6] Acet	257641
H3	Hht1 / Hht2	FQK**STELLIR**	[3] Acetyl (K)	2953502
H3	Hht1 / Hht2	RFQK**STELLIR**	[4] Acetyl (K)	32597
H3	Hht1 / Hht2	EIAQDFKTDLR		502177
H3	Hht1 / Hht2	EIAQDFK		1122166
H3	Hht1 / Hht2	**STELLIR**		9115570
		STELLIR		**12101669**

Abundance and stoichiometry of *A. gossypii* histones

The goal of this histone quantification was to compare the theoretical ratios of the four core histones H2A, H2B, H3, and H4 (1:1:1:1) with the experimentally determined ratios. In addition, we also discussed the abundance found for the H2A like histone Htz1 which replaces H2A in a small subset of nucleosomes. Thus, the abundance of H2A plus Htz1 has to be considered when the experimentally determined ratios are discussed. We observed deviations of different degrees depending on the quantification method, as summarized in Table2-14. Using the All-method, we found very similar abundances for H2A and H2B but a 16fold excess of H4 over H3. Compared to H2A/H2B, histone H3 was threefold underrepresented and histone H4 was sixfold overrepresented. This deviation can be partly explained by the low number of identified H3-specific peptides which very likely have a low ionization potential. Also, several of the H2A and H2B specific peptides most likely have a low ionization potential. The H2A like histone Htz1 was found with 2% abundance compared to H4, and it was ninefold less abundant than H2A. A similar picture emerged, when the relative abundances were determined with the Top3-method. The only difference is a twofold higher abundance for H2B compared to H2A. By using these two methods, the three histones H2A, H2B, and H3 are underrepresented three to sixteenfold compared to H4.

Table2-12: MASCOT output of the detected peptides and measured abundances of the two *A. gossypii* H4 histones. The table list for each peptide the sequence, the modification, the normalized abundance, the score, and the fraction of the OffGel-electrophoresis in which the peptide was detected. The validation "TRUE" or "FALSE" indicates whether the peptide was included or excluded for the automated calculation of the protein abundance.

Accession	common name	Sequence	Modifications	0h Normalized abundance	Use in quantitation	Score	Fraction
ADL201W	Hhf2	GLGKGGAKR	[8] Acetyl (K)	0.554372786	FALSE	21.72	12
ADL201W	Hhf2	DNIQGITKPAIR		18163564.72	TRUE	38.05	12
ADL201W	Hhf2	DNIQGITKPAIR		27218740.48	TRUE	60.36	12
ADL201W	Hhf2	TVTSLDVVYALKR		150415.8939	TRUE	43.59	12
ADL201W	Hhf2	TVTSLDVVYALKR		365125.0952	TRUE	81.59	12
ADL201W	Hhf2	KTVTSLDVVYALK		70170.93553	TRUE	112.12	12
ADL201W	Hhf2	KTVTSLDVVYALK		17323.24832	TRUE	40.79	12
ADL201W	Hhf2	KTVTSLDVVYALK		607323.0705	TRUE	90.16	12
ADL201W	Hhf2	TVTSLDVVYALK		2527.111223	TRUE	33.33	12
ADL201W	Hhf2	TVTSLDVVYALKR		2613874.125	TRUE	40.66	11
ADL201W	Hhf2	TVTSLDVVYALKR		1268501.205	TRUE	76.89	11
ADL201W	Hhf2	KTVTSLDVVYALK		223249.3068	TRUE	46.03	11
ADL201W	Hhf2	KTVTSLDVVYALK		467174.8886	TRUE	59.1	11
ADL201W	Hhf2	KTVTSLDVVYALK		92258.09404	TRUE	28.88	11
ADL201W	Hhf2	KTVTSLDVVYALK		147862.4177	TRUE	84.65	11
ADL201W	Hhf2	DNIQGITKPAIR		5891.748321	TRUE	25.23	11
ADL201W	Hhf2	TVTSLDVVYALKR		139430.7344	TRUE	24.27	10
ADL201W	Hhf2	TVTSLDVVYALKR		34891.03352	TRUE	65.56	10
ADL201W	Hhf2	TVTSLDVVYALK		57406.9207	TRUE	57.76	10
ADL201W	Hhf2	AVLKSFLESVIRDAVTYTEHAK		321.4247254	TRUE	40.34	8
ADL201W	Hhf2	DNIQGITK		88074.90302	TRUE	31	7
ADL201W	Hhf2	ISGLIYEDVR		1197334.866	FALSE	17.3	6
ADL201W	Hhf2	TVTSLDVVYALK		220433.6236	TRUE	49.71	6
ADL201W	Hhf2	ISGLIYEDVRAVLK		1306.682779	TRUE	37.05	6
ADL201W	Hhf2	SFLESVIR		19904049.93	TRUE	36.94	5
ADL201W	Hhf2	SFLESVIR		37387.0842	TRUE	22.56	5
ADL201W	Hhf2	TVTSLDVVYALK		9224587.447	TRUE	62.09	5
ADL201W	Hhf2	SFLESVIRDAVTYTEHAK		33885.1959	TRUE	48.41	5
ADL201W	Hhf2	SFLESVIRDAVTYTEHAK		2780.376586	TRUE	19.02	5
ADL201W	Hhf2	DAVTYTEHAK		1365.001913	TRUE	40.79	4
ADL201W	Hhf2	DAVTYTEHAK		8541312.506	TRUE	19.54	4
ADL201W	Hhf2	DAVTYTEHAK		3439.226152	TRUE	28.57	4
ADL201W	Hhf2	TVTSLDVVYALK		1098765.555	TRUE	47.39	4
ADL201W	Hhf2	TVTSLDVVYALK		4557532.401	TRUE	27.47	4
ADL201W	Hhf2	DNIQGITK		534687.2853	TRUE	33.05	4
ADL201W	Hhf2	SFLESVIR		205064.088	TRUE	15.19	4
ADL201W	Hhf2	SFLESVIRDAVTYTEHAK		40231.53879	TRUE	51.2	4
ADL201W	Hhf2	SFLESVIRDAVTYTEHAK		2705.549687	TRUE	30.56	4
ADL201W	Hhf2	DNIQGITK		686571.9079	TRUE	28.58	3
ADL201W	Hhf2	SFLESVIR		221719.2281	TRUE	29.31	9
ADL201W	Hhf2	ISGLIYEDVR		226158.6524	TRUE	48.08	9
ADL201W	Hhf2	ISGLIYEDVR		43466315.26	FALSE	55.87	2
ADL201W	Hhf2	ISGLIYEDVR		102408.058	TRUE	18.04	2

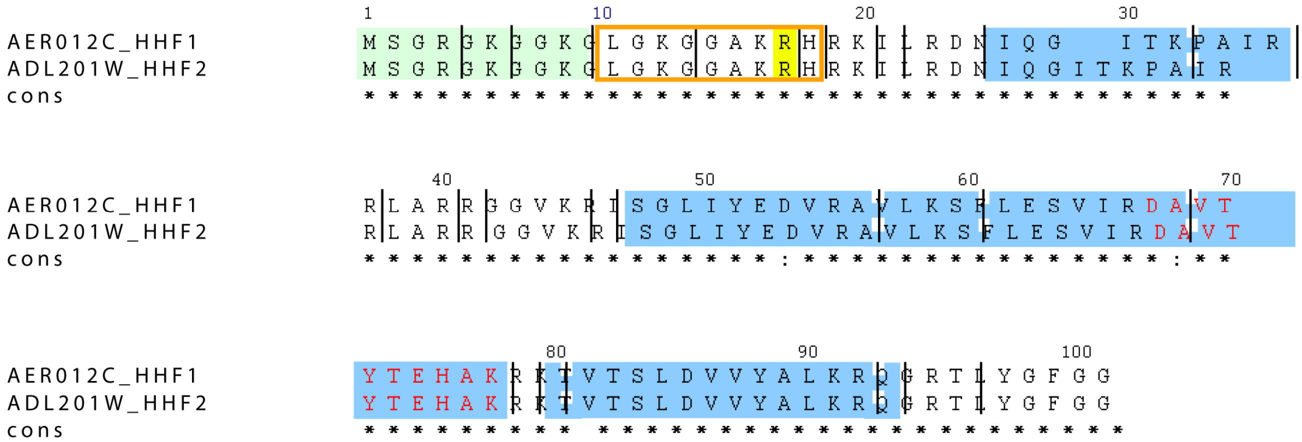


Figure2-13: Sequence alignment of the two H4 histones and experimentally identified peptides. Detected peptides are marked in blue. Lysines highlighted in yellow can be acetylated resulting in detected peptides marked by orange brackets. Peptides with a score less than 15 were not used for protein quantification and marked in light green. The AQUA peptide used in subsequent experiments ist marked in red letters (see Chapter II-C).

Table2-13: Compilation of all highly abundant peptides of histones Hhf1 and Hhf2. Color codes facilitate to visualize the same sequence in completely or partially cleaved peptides. At the bottom of the table the four most abundant peptides are listed.

histone	common name	Sequence	Modifications	0h Normalized abundance
H4	Hhf1 / Hhf2	AVLKSFLESVIRDAVTYTEHAK		321
H4	Hhf1 / Hhf2	DAVTYTEHAK		8546117
H4	Hhf1 / Hhf2	**DNIQGITK**		1309334
H4	Hhf1 / Hhf2	**DNIQGITK**PAIR		45388197
H4	Hhf1 / Hhf2	GLGKGGAKR	[8] Acetyl (K)	1
H4	Hhf1 / Hhf2	**ISGLIYEDVR**		44992217
H4	Hhf1 / Hhf2	**ISGLIYEDVR**AVLK		1307
H4	Hhf1 / Hhf2	KTVTSLDVVYALK		1625362
H4	Hhf1 / Hhf2	**SFLESVIR**		20368220
H4	Hhf1 / Hhf2	**SFLESVIR**DAVTYTEHAK		79603
H4	Hhf1 / Hhf2	**TVTSLDVVYALK**		15161253
H4	Hhf1 / Hhf2	**TVTSLDVVYALK**R		4572238
		DNIQGITK		**46697531**
		ISGLIYEDVR		**44993524**
		SFLESVIR		**20447823**
		TVTSLDVVYALK		**19733491**

We therefore tried a third method, relying only on the most abundant peptide found for each histone. These peptides and their relative abundances are listed in the right part of Table2-14. Now, only the two histones H2A and H3 were three to fourfold under epresented compared to H2B and

H4, both of which show similar abundances. Interestingly, based on this method the H2A-like histone Htz1 was found to be only fourfold less abundant than H2A. In summary, the result obtained with the most abundant peptide method deviates much less from the theoretical 1:1:1:1 ratio compared to the other methods. As discussed above, the relative abundance of H2A can also be estimated as 19.5 million (not as 12.5 million) based on the high abundance of a peptide which is present in Hta1, Hta2 and Htz1, and the fairly reliable estimate of the Htz1 abundance, which is 3.24 million This would increase the relative abundance of H2A by 50 % which would mean that H2A is replaced by Htz1 in one of seven nucleosomes in spore chromatin.

Histone acetylation

Chromatin remodeling by modification of the histones is essential for modulation of transcription activity of genes. Histone modifications have so far not been investigated in fungal spores. We therefore compared histone acetylation sites of N-terminal histone tails in growing *S. cerevisiae* cells (Li et al., 2007) with histone acetylations in *A. gossypii* spore chromatin. All lysines at the amino-terminus of histones are conserved in both fungi, except for lysines K17 and K18 which are deleted in one of the H2B copies of *A. gossypii*. The high conservation of lysine sites is a prerequisite for analyzing functional differences in histone acetylations, for example the differences between silenced spore chromatin and active chromatin in related fungi as summarized in Figure2-14. Interestingly, we detected the same acetylation sites only for the H2A like histone Htz1. The histone pair H2A and H2B was higher acetylated in *A. gossypii* spores than in growing *S. cerevisiae* cells. In contrast, the histone pair H3/H4 was less acetylated in *A. gossypii* spore chromatin. These data show encouraging differences between transcriptionally silenced and active chromatin. As a next step the lysine acetylation pattern in growing *A. gossypii* hyphae needs to be investigated to see whether it significantly differs from the pattern found in spores and whether it is similar to the pattern observed in growing *S. cerevisiae* cells.

Table2-14: Histone abundances determined with three different methods. Numbers in the abundance columns were obtained by the All-Method, the Top3-Method, and the Most Abundant Peptide-Method. The percent columns list the abundances of the respective histones relative to histone H4 (100%). See text for further details.

Histone	0h relative abundance all peptides	percent relativ to H4 (100%)	0h relative abundance top3 peptides	percent relativ to H4 (100%)	most abundant specific peptide	0h relative abundance most abundant specific peptides	percent relativ to H4 (100%)
Htz1	27.6	2	46.3	2	DGGPLGSQSHSAR	307	7
Hta1 / Hta2	270.0	17	505.3	19	LLGNVTIAQGGVLPNIHANLLPK	1246	27
Htb1 / Htb2	273.6	17	1002.4	37	LILPGELAK	5109	109
H3 (2 copies)	97.3	6	152.6	6	STELLIR	1210	26
H4 (2 copies)	1623.0	100	2697.3	100	DNIQGITK	4670	100

Discussion

Peptide ions vary significantly in their ability to be detected by mass spectrometry instruments, since biochemical sequence properties can affect peptide ionization (Steen and Pandey, 2002). The differences in ionization affect the measurements of the abundance of peptides by several orders of magnitude and therefore play a crucial role in computing protein abundances. We used two different ways to deal with this methodological problem to determine relative protein abundances. In the All-method the sum of the measured abundances of all peptides of each protein (including peptides from partial digestions and modified peptides) is divided by the number of theoretically detectable peptides. The obtained numbers are taken as a measure for relative protein abundances. If only a few peptides from the theoretically detectable peptides of a protein are detected, the All-method underestimates protein abundances. Therefore the Top3 method was introduced in which the mean of the 3 most abundant peptides is regarded as a reliable estimate for the relative abundance of proteins.

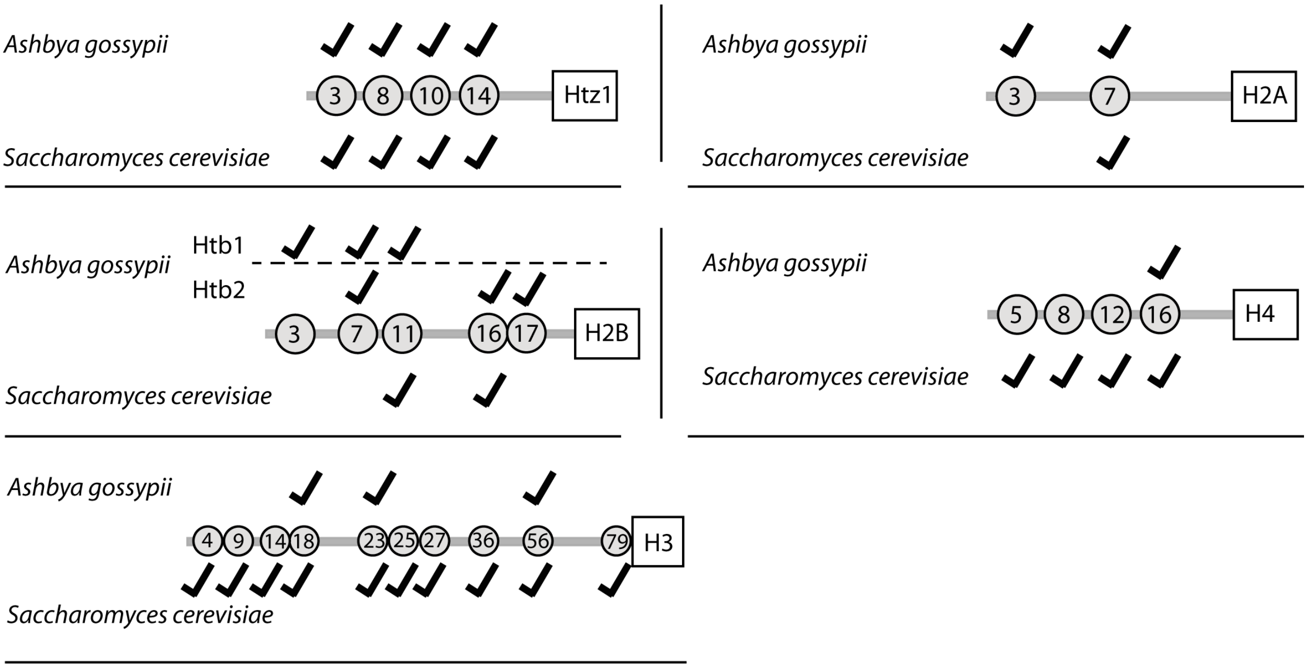


Figure2-14: Histone acetylation map of *A. gossypii* and *S. cerevisiae*. Numbers indicate the sequence position of the modified lysines in the mature histones, which often lack the amino-terminal methionine. The positions of the lysines are conserved in *A. gossypii* and *S. cerevisiae*. For *A. gossypii* the detected acetylations of lysines are indicated with tick marks. The indicated positions of acetylated lysines in growing *S. cerevisiae* cells were taken from a review (Li et al. 2007).

The ratios of the relative abundances of H2A plus Htz1, H2B, H3, and H4 obtained by the different methods are summarized in Table2-14. With H4 taken as 1.0 these ratios are 0.19:0.17:0.06:1.0 for the All-method and 0.21:0.37:0.06:1.0 for the Top3-method. The number of nucleosomes in which H2A is replaced by Htz1 is about 1 in 10 comparing the abundances of Htz1 and H2A or 1 in 50 comparing the abundances of Htz1 and H4 as calculated by both methods.

The ratio of histone abundances calculated based on the most abundant peptide is 0.35:1.09:0.26:1.0. A outlined above, this ratio is most likely 0.49:1.09:0.26:1.0 based on the most abundant peptide HLQLAIR found in H2A and Htz1. According to this method the number of nucleosomes in which H2A is replaces by Htz1 is 1 in 5 to 1 in 6 comparing the abundances of Htz1 and H2A or 1 in 13 to 1 in 14 comparing the abundances of Htz1 and H4. The relatively high occurrence of Htz1 in spore chromatin makes sense because Htz1 plays a role in transcriptional silencing and probably also chromatin condensation (Meneghini et al., 2003). Histone H3 is still fourfold underrepresented with this improved approach to estimate histone abundances (Table2-14). As already mentioned above, the low coverage of detected peptides precludes a realistic estimate of H3 abundance. When AQUA-peptides specific for histones H3 and H4 are used for absolute quantifications (see next section) equal amounts of histone H3 and H4 were found.

The low acetylation of histones H3 and H4 found in *A. gossypii* spores is in agreement with published data of transcriptionally active chromatin of *S. cerevisiae* which indicated that increased acetylation of lysines in histone H3 and H4 leads to increased gene activity (Workman and Kingston, 1998). As was shown later, the acetylation of histones H3 and H4 is necessary to improve promoter activities (Pokholok et al., 2005). The low level of acetylation is therefore expected for silenced chromatin in *A. gossypii* spores.

Part C: Absolute protein copy numbers per spore

Introduction

Spores are in a quiescent state meaning that all metabolic and cell biological activities are silenced. We assume that the protein composition described in the previous section is very similar in all spores because these spores were isolated from mycelia growing on plates with identical amounts of nutrients. Therefore, the condition for sporulation (nutrient depletion) was as similar as possible for all mycelial colonies collected for spore isolations. Knowing the relative abundance of all proteins in the spore proteome we also wanted to determine the copy number per spore for each detected protein. In order to reach this goal the amount of protein in a single *A. gossypii* spore has to be known which will allow the translation of the relative abundances into copy numbers per spore. Two independent approaches were tried. In the first approach the weight of a single spore was estimated via the average spore volume, and the protein content of a spore was determined as percent of the spore biomass. In the second approach the number of spores was estimated which was needed to isolate a specific amount of spore proteins. This second approach allowed to directly determine the amount of protein in a single spore without knowing the weight of a single spore and its protein content.

The two attempts to determine the amount of protein in a single spore will be described in the first two sections including the steps for calculating copy numbers of proteins in a single spore based on their relative abundances (chapter B). The two methods used for translating MS intensities of peptides into relative abundances of proteins are well accepted but lack independent controls, for example by adding heavy isotope-labeled reference peptides of known concentration to the peptide mixture prior to LC-MS analysis. Such reference peptides were synthesized and used as internal calibration standards for the high-resolution MS run of the tryptic peptides from spore proteins. The last two sections of this spore proteome Chapter will therefore focus on the use of exactly known quantities of heavy isotope-labeled reference peptides (AQUA peptides). First the steps for absolute quantification of spore proteins represented by a reference peptide, and the determination of their copy numbers will be documented. The final section will represent the generation of an AQUA peptides-based regression line and its use to directly translate relative protein abundances into copy numbers per spore.

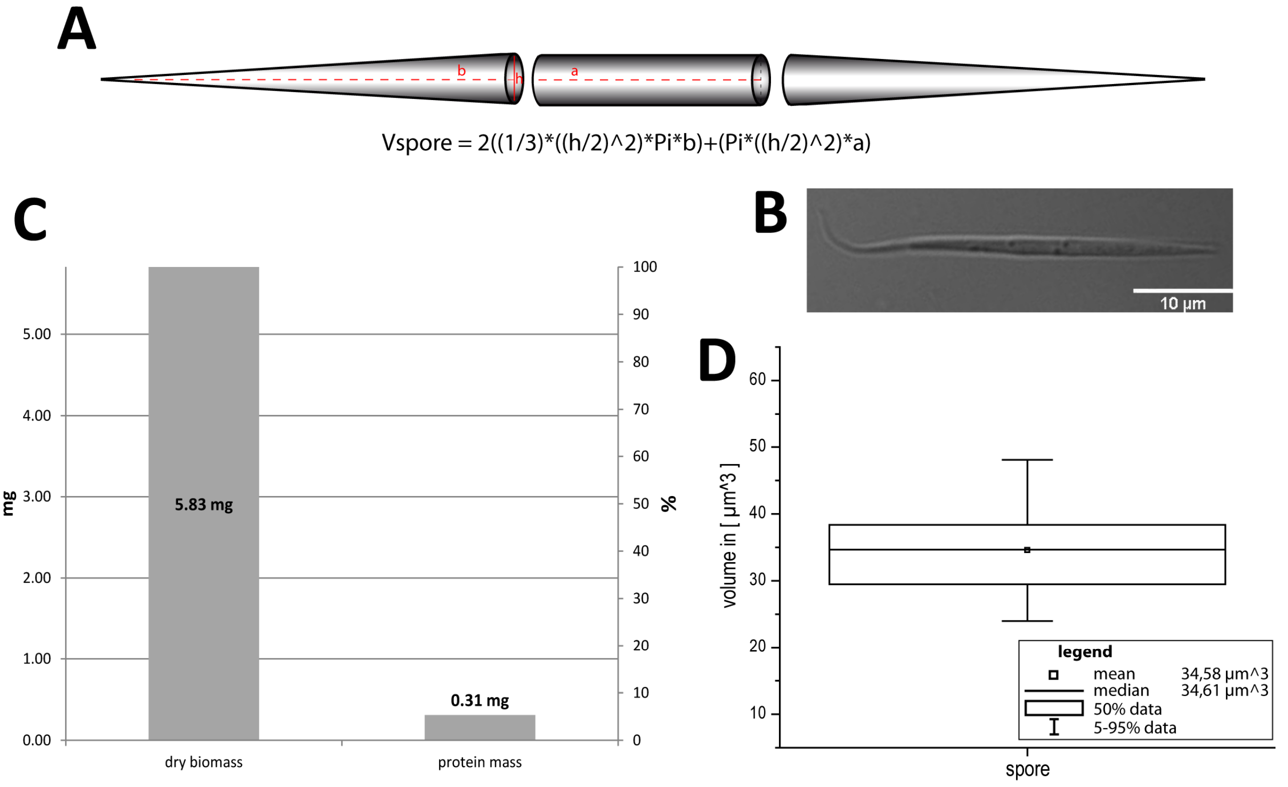
Results

The spore volume approach to define numbers of proteins per spore

Because we cannot weigh single spores we started to measure the volume of spores and then estimated their density to know the average weight of a single spore. As shown in the scheme of Figure2-15A the needle-shaped spores were subdivided into three zones, a central cylinder and two cones. The scheme is an acceptable approximation of the shape of an *A. gossypii* spore shown in Figure2-15B. The lengths and the diameters of the cylinder and the cones were measured for 56 spores using the image-processing program ImageJ. The number of pixels for each dimension was counted in the 16bit unprocessed images and was converted into µm knowing the resolution of the Photometrics camera (0.0645 µm/pixel). The average total length of a spore is 34 µm and the average diameter 1.96µm. The volumes were calculated for each spore using the formula written underneath the scheme. In Figure2-15D the distribution of calculated spore volumes is plotted. 50% of the volumes are between 29 μm^3 and 37 μm^3, and a mean of 34.6 μm^3 was calculated for a spore volume.

Next the protein concentration in spores was determined. Proteins were extracted from a known amount of lyophilized spores by using the maximal output of a FastPrep Cell Disrupter and 0.5 mm zirconia/silica beads followed by sonification as described in Materials and Methods. From 5.83mg spores 0.31mg protein could be extracted as measured with a Bradford protein assay (Figure2-15C). This low amount of 5.63% protein in the spore biomass suggests an unusually high content of glucan, chitin and chitosan in *A. gossypii* spores. In order to estimate the amount of protein for a single spore an equal distribution of the spore density was assumed which means that 5.63% of the spore volume (1.94 μm^3) consists of proteins. As outlined in Figure2-15E this small volume contains approximately 2.5 pg protein based on the assumption that the density is 1.29 g/ml. Using CsCl density centrifugations the density of *A. gossypii* spores was recently estimated as 1.37 g/ml (H.-P. Schmitz personal communication). This is fairly close to the value used in Figure2-15 and therefore the calculations presented in this section are still based on 2.5 pg as average weight of a spore.

To calculate the copy numbers per spore for each protein we converted the relative protein abundances (Table2-2 Appendix 1) into relative protein amounts by multiplication with the molecular weight of each protein. The sum was taken as 100% and the amount of each protein was calculated as fraction of 2.5 pg. The actual copy numbers per spore were calculated by dividing the amounts per spore with the molecular weight and the atomic mass unit 1.6605×10^{-24}. The copy numbers for the top 20 proteins are presented in Table2-15 discussed below. For the complete Table see Appendix 5



E

request	known values	calculation	result
protein volume in a spore	spore volume*protein % in spores/100	34.58*5.6/100	1.94 µm^3
volume in femto liter	1µm^3 = 1 femto liter	1.94*1	1.94 femto liter
in liter	1 femto liter = 10^-15l	1.94*10^-15	1.94*10^-15 liter
density spore	p(density)= mass/volume ((%protein*MW)+(%H2O*MW))/(100*MWH2O)	((5.6*110)+(94.4*18))/(18*100)	1.29
weight	density 1.29 * 1kg=1l	1.94*10^-15 * 1.29	2.5E-15 kg
in pico gram	1kg=10^15 pico gram	2.5*10^-15 * 10^15	2.5 pg
			2.5 pg protein per spore

Figure2-15: Protein concentration and volume of *A. gossypii* spores. **(A)** Schematic projection of an *A. gossypii* spore. The lengths of the dashed red lines were measured in images of 56 spores and used to calculate the spore volume according to the listed formula (volume of the two cones plus volume of the cylinder). **(B)** Microscopic DIC image of a spore taken with a 63fold magnifying objective. Calibration bar 10 µm. **(C)** Protein mass in correlation with the biomass, measured in mg (y-1 axes) and specified in percent at the y-2 axes. Spores contain 5,6% protein. **(D)** Box plot of the distribution of 56 calculated spore volumes. The mean of the spore volume of the 56 spores is 34,5 cubic micrometer. **(E)** Computing steps for estimating the amount of protein per spore. The protein volume is 1.94×10^{-15} liter (1 μm^3 is equal to 1 femto liter) per spore. Water has a density of 1 and 1l is equivalent to 1 kg. Because the density of *A. gossypii* spores is unknown, the density of the protein volume was estimated as 1.29 (extrapolated from the protein water proportion and the molecular weight). According to this estimate approximately 2.5 pg protein are present in one *A. gossypii* spore.

Table2-15: Copy numbers per spore of the most abundant spore proteins based on relative abundances determined with the All –method and determined 2.5 to 2.8 pg of protein per spore.

accession	common name	cps [All method; %]
AAL128C	Hsp26	4373420
ADL370C	Tef1, Tef2	1308640
AGL321W	Sod1	1057296
ADL154C	NOHBY412	1039105
AFR694W	YLR179C, Tfs1	728931
AER312W	Tsa1, Tsa2	674749
ABL156C	Fmp45	664228
ABR068C	Fba1	478315
ADL032W	Aco1	477364
ABR028C	Cwp1	473801
ABR222W	Act1	431184
AFR020W	Ach1	417070
AGL177C	Cpr1	403472
AER428W	Om45	348844
ACL198W	Ilv5	311512
AER401W	Ypr1, Gcy1	311489
AAR004C	Cit1, Cit2	294888
AER031C	Tdh3	272662
AER184W	Pet9, Aac3	251121
AFR279C	Hxk2, Hxk1	248318

The first steps of these conversions are presented in Table2-16 for the top 20 proteins. Columns 1, 3, and 4 refer to the names of the proteins, their relative abundance and their molecular weight, respectively. Column 6 lists the protein amounts expressed as percent of the spore proteome. Column 2 and 5 show the ranks of the most prevalent spore proteins according to their amounts or abundance. Proteins with a relatively high molecular weight are ranked higher in column 2 compared to column 5.

Table2-16: The top 20 proteins ranked by their weight percent. Percentages are calculated with the molecular weight and the total protein abundance. The measured abundances are divided by 4 times the decimal power to avoid too long numbers. Column 2 shows the rank based on the abundance.

common name	rank by abundance	0h relative abundance All-method (10^{-4})	protein size [MW]	rank by percent	% in total extract
Hsp26	1	172894	20199	1	5.87
Tef1, Tef2	2	51734	49876	2	4.34
Aco1	9	18872	84557	3	2.68
Cwp1	10	18731	70107	4	2.21
Fmp45	7	26259	47034	5	2.08
Ach1	12	16488	57754	6	1.60
NOHBY412	4	41079	21035	7	1.45
Fba1	8	18909	39319	8	1.25
Om45	14	13791	53827	9	1.25
Act1	11	17046	41688	10	1.19
Sod1	3	41798	15856	11	1.11
YLR179C, Tfs1	5	28817	22333	12	1.08
Cit1, Cit2	17	11658	52946	13	1.04
Tsa1, Tsa2	6	26675	21486	14	0.96
Atp1	22	9161	58887	15	0.91
Ilv5	15	12315	43687	16	0.90
Hxk2, Hxk1	20	9817	53943	17	0.89
Sdh1	31	7509	68710	18	0.87
Uga2	21	9609	53624	19	0.87
Pdc1, Pdc5	29	7976	64130	20	0.86

With 5.8% of the total proteome the small heat shock protein Hsp26 is by far the most prominent spore protein. Each spore carries about 4 million copies. It most likely acts as a molecular chaperone (Petko and Lindquist, 1986) and could play an important role in stabilizing protein structures during the long lasting spore stage to allow efficient spore germination. This protein is indeed spore-specific because its abundance decreases substantially during development (Chapters C and D). From the remaining 3,803 identified spore proteins (error probability of 1%) 3,731 proteins have a copy number of 1 or more per spore. Only 72 of the proteins have a lower calculated copy number than 1. The lowest one is Mak3 with 0.0004cps which means that it is present in only 1 of 2500 spores. The yeast homologue of this protein plays a role in replication of double-stranded RNA plasmids (Wickner and Leibowitz, 1976). Since *A. gossypii* does not carry circular double –stranded RNA it is very likely a false positive.

The circular diagrams of Figure2-16 show that rather small numbers of proteins dominate the spore proteome with respect to the protein amounts. For example, the 20 proteins listed in Table2-16 represent already one third of the spore proteome. The top 200 proteins represent three quarters and the top 1,000 proteins over 95% of the proteome leaving 5% of the total amount for the remaining 2,700 spore proteins.

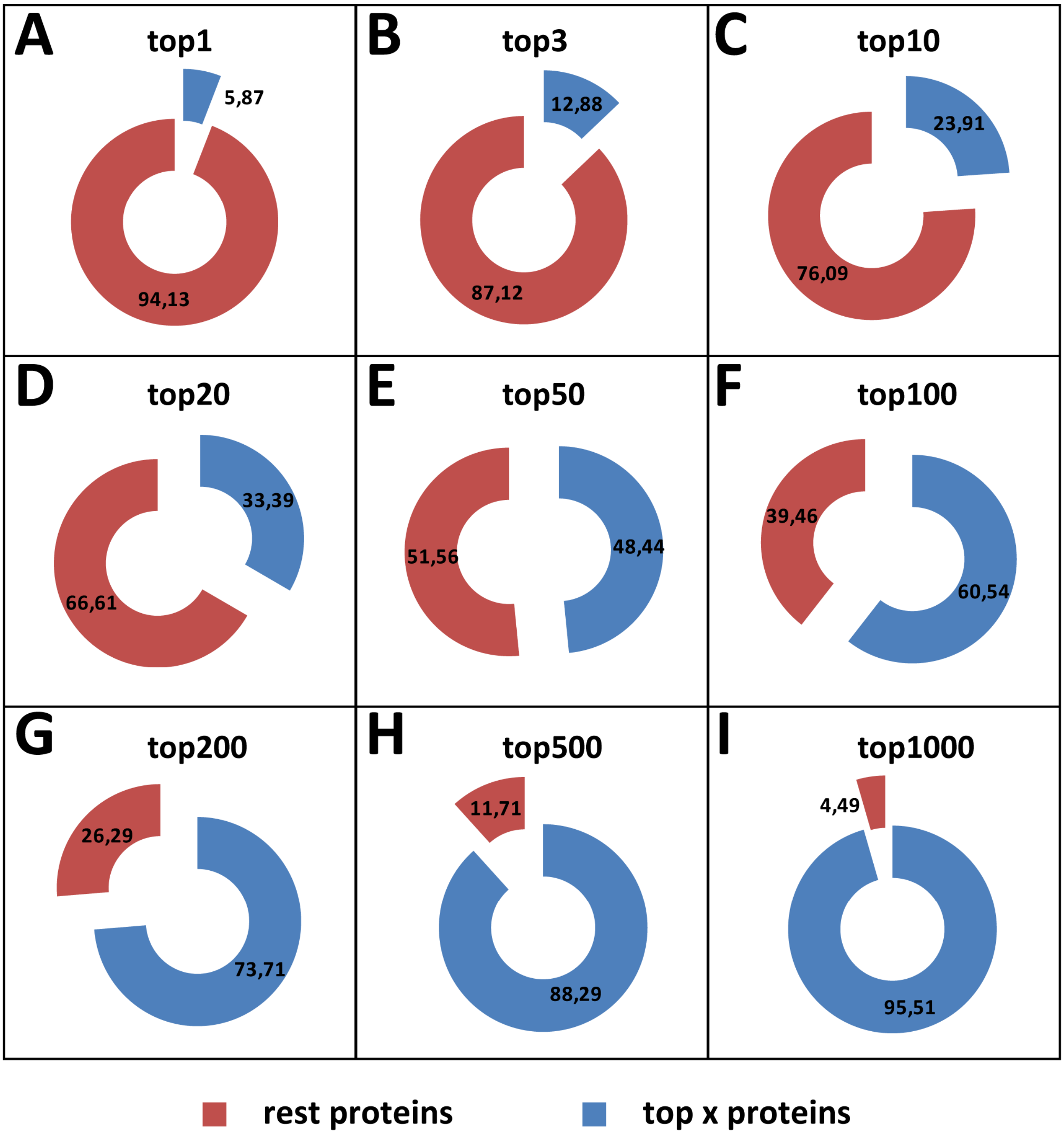
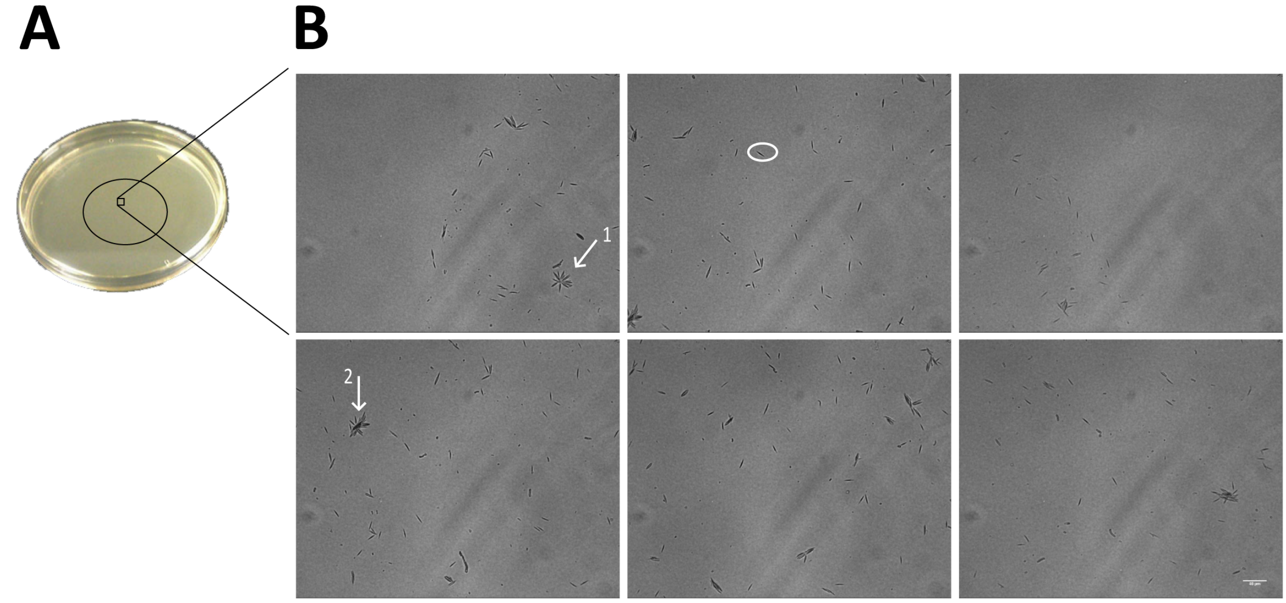


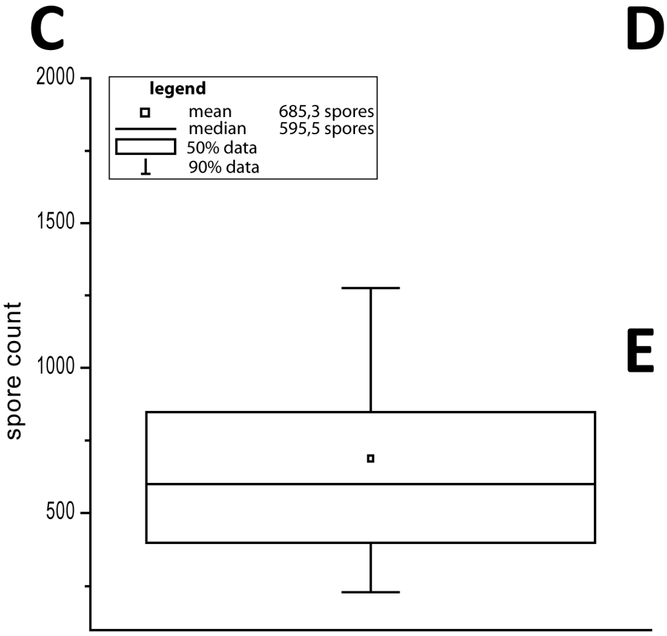
Figure2-16: The spore proteome is dominated by a relatively low number of all spore proteins. The circular diagrams in the nine panels document the weight percent of the most abundant protein **(A)** to the 1,000 most abundant proteins **(I)** with respect to the amount of all 3,800 detected spore proteins. The identities of the top 20 spore proteins are listed in table2-16.

The spore titer approach to define numbers of proteins per spore

We tried to determine the titer of a fresh spore preparation and measured the amount of protein which could be isolated from these spores (Figure2-17). Due to their hydrophobic surface *A. gossypii* spores have the tendency to frequently form aggregates. Therefore serial dilutions and counting of spores is not an easy task. We performed several 1:10 dilution steps with rigorous vortexing prior to each dilution step. From the lowest dilution 5 µl aliquots each were spread in a marked area in the center of 20 AFM agar plates (Figure2-17A). For each marked area 40 to 50 microscopic images were taken and aligned prior to counting the needle-shaped spores. Figure2-17B shows some images of spores from one area. Single spores are clearly identifiable with this magnification. The aggregate labeled as 1 consist of 12 spores. The aggregate labeled as 2 contains most likely 16 spores. In several images larger aggregates were seen and the numbers of spores could only be estimated as 50 or 100. This at least partly explains the variation in the number of spores counted in each area with an average of 685 spores (Figure2-17C). The calculation for the total number of spores used to extract 2 mg of protein is documented in Figure2-17D and E. Also the loss of some material during the sample preparations was taken into account. For example we used about 2% of the 2 mg protein for the SDS-PAGE control. The total protein extract was made from 0.9×10^9 to 1×10^9 spores which means that on average 2.1 pg protein were isolated from each spore. This is slightly lower than the 2.5 pg protein per spore estimated with the spore volume/density approach. It is possible that the number of spores was overestimated in the larger aggregates and therefore the titer of the spore preparation was probably slightly lower. Until a more reliable spore counting method becomes available, we rely on 2.5 pg as average protein content in one spore.

The steps for the translation of relative abundance to the mass percent of each protein, the division of the 2.5 pg protein (100%) into the absolute mass of each protein in one spore, and the conversion of the absolute mass into copy number per spore was already described in the previous section. The obtained numbers are only estimates because they rely on the translation of the measured MS peptide intensities into estimates of relative abundance of proteins, and the two methods employed are not optimal as discussed in the histone section of Chapter2-B. The use of precisely known concentrations of heavy isotope-labeled reference peptides added to the peptide mixture can lead to adjustments of the relative protein abundances and with that improved copy numbers per spore as described in the following two sections.





D

used	spores
spores per Petri dish (5µl]	685.3
spores per µl	137.06
dilution to count(1:1000)	137060
dilution to store/fix (10µl in 1300µl)	17817800
10µl used from 600µl total volume	1069068000
spore count	1.07E+09

E

	used µl	in %	used spores	spores
protein amount definition	10	1	10690680	1058377320
SDS gel predigest	10	1	10583773	1047793547
SDS gel supernatet fällung	20	1.4	14669110	1033124437
SDS gel +trypsin	30	1.4	14463742	1018660695
SDS gel 1/4h digest	30	1.4	14261250	1004399445
SDS gel 1/2h digest	30	1.5	15065992	989333454
SDS gel 1h digest	30	1.5	14840002	974493452
SDS gel o/N digest	30	1.5	14617402	959876050
MS test	50	2.3	22077149	937798901
				$9.38*10^8$

Figure2-17: Number of *A. gossypii* spores needed to obtain 2 mg protein. **(A)** Agar petri dish with a marked area for spreading spores from 5 µl of the highest spore dilution. **(B)** Microscopic images of six areas with spread spores. Scale bar 40 µm. The circle marks a single spore. Arrows point to two bundles of hydrophobic spores with 12 spores (1) and estimated 16 spores (2). 40 to 50 images were aligned to count all spores spread on one agar plate. **(C)** Box plot showing the distribution of counted spores from 20 agar plates. The average is 683 spores per 5 µl. **(D)** Calculation of the original number of spores present in 600 µl concentrated spores. **(E)** Steps at which different aliquots were withdrawn from the trypsin digestion of the 2 mg spore protein. The numbers refer to the respective equivalents in spore numbers. AQUA peptides (2 pmol or 20 pmol) were spiked into the peptide mixture at the end of the trypsin digestion as described in the text.

Table2-17: List of AQUA peptides and the corresponding proteins they were selected from. The listed peptides with isotope-labeled C-terminal arginine or lysine were added either as 20pmol (10x) or 2pmol (1x) to the spore peptide mixture. The decision was based on estimated abundances extrapolated from the peptide coverage found in pre-runs (column 5 and 8).

accession	common name	number of unique peptides	total number of peptides	per cent count	peptide	length	spike	note
ABR222W	Act1	62	234	65.2	AGFAGDDAPR	10	10x	
AER085C	Fas1	250	736	62.6	LLADDEQPTALPK	13	10x	
AEL329W	Pil1	40	86	58.9	IEVLEQELVR	10	10x	
AFR441C	Tub1, Tub3	38	100	49.6	ELLDDILDR	9	10x	
AFR320W	Lsb1, Pin3	13	15	46.3	QQSEQIESNLPNPNEALR	18	10x	
ADR058C	Cdc28	16	21	46.1	LGDFGLAR	8	10x	
AER359W	Chc1	104	150	43.8	TGQFQEVQR	9	10x	
AFR140C	Rvs167	18	20	42.5	YGSEASPSPTATFR	14	10x	
AGL237C	Abp1	33	51	42.2	LLLVGWCPDSAPLK	14	10x	no detection
AFL181C	Erg2	11	14	38.7	DALAAHYGEEYINR	14	10x	
AEL241W	Bem1	23	53	37.9	TPPIGSPQAAGSQLK	15	10x	
ABR028C	Cwp1	73	437	37.5	LSAAGEALGQAAGR	14	10x	
AGL093W	Cdc42	10	18	34	LRPITPEQGEK	11	10x	
AEL025W	Cdc14	14	33	30.4	VYLDNTIEFLK	11	10x	
AAL181C	Net1, Tof2	34	67	29.9	SILPPPPPQSPPIR	14	10x	
ADL201W	Hhf2	5	22	29.1	DAVTYTEHAK	10	10x	
AGL169C	Nbp2	9	9	28.6	HGQGWLVAEDSAR	13	10x	
ADR018C	Pan1	32	51	26.4	SAVRPGATTITGEDCK	16	10x	no detection
AGR285W	Las17	16	24	23.5	LYIAYPNPK	9	10x	
AFR464W	Rsr1	5	9	20.8	SNVDEVFIDLVR	12	10x	
AGR306C	Bbc1	15	15	19.7	EAEEEDEEEDNAGLK	15	1x	no synthesis
AEL306C	Myo5, Myo3	24	26	17.6	HGNQFEQLR	9	1x	
AEL006W	Utp20	45	49	15.9	ITDNFILPDR	10	1x	
AGL293C	Boi2, Boi1	14	18	14.9	LSEFYSPGR	9	1x	
AGR043W	Num1	49	46	14.4	ELNDENDQLNSK	12	1x	
ABR008C	Hse1	5	4	13.8	LQEDQDLEEALK	12	1x	
ADL022C	Spa2, Sph1	37	39	13.3	TQVISTGVSPAR	12	1x	
AEL017W	Lsb3, Ysc84	5	4	12.3	AILSGSVSPPPGVDPLFR	18	1x	
AFR702W	Bik1	6	3	11.5	LPNESLPLFDGNK	13	1x	no detection
AFR100W	Exo70	5	3	10.6	ILNPNLEPEVGAPHK	15	1x	
AFR100W	Exo70	5	3	10.6	DLASNLLDQVFVDSSGK	17	1x	
ADR033W	Dbf2, Dbf20	5	5	10.4	NLEFPEFK	8	1x	
AGR095W	Rax2	14	14	10.4	INEDEMLK	8	1x	
ADL202C	Hht2	2	4	10.3	STELLIR	7	1x	
ADL038W	Cdc25, Sdc25	14	12	10.2	SPFTELLPVDPLQR	14	1x	
AGL364C	Bnr1	8	15	10.1	EDGDLLSR	8	1x	
AFR669W	Bni1	16	9	8.8	HEHEVELR	8	1x	
ACR258W	Dyn1	32	23	8	ILEESQLSPK	10	1x	
AEL090C	Inp54	5	2	7	LNSGVEDTPEVPVDR	15	1x	
ACR145W	Kip2	3	1	5	GLLEVENSILK	11	1x	

Absolute protein quantification with AQUA peptides

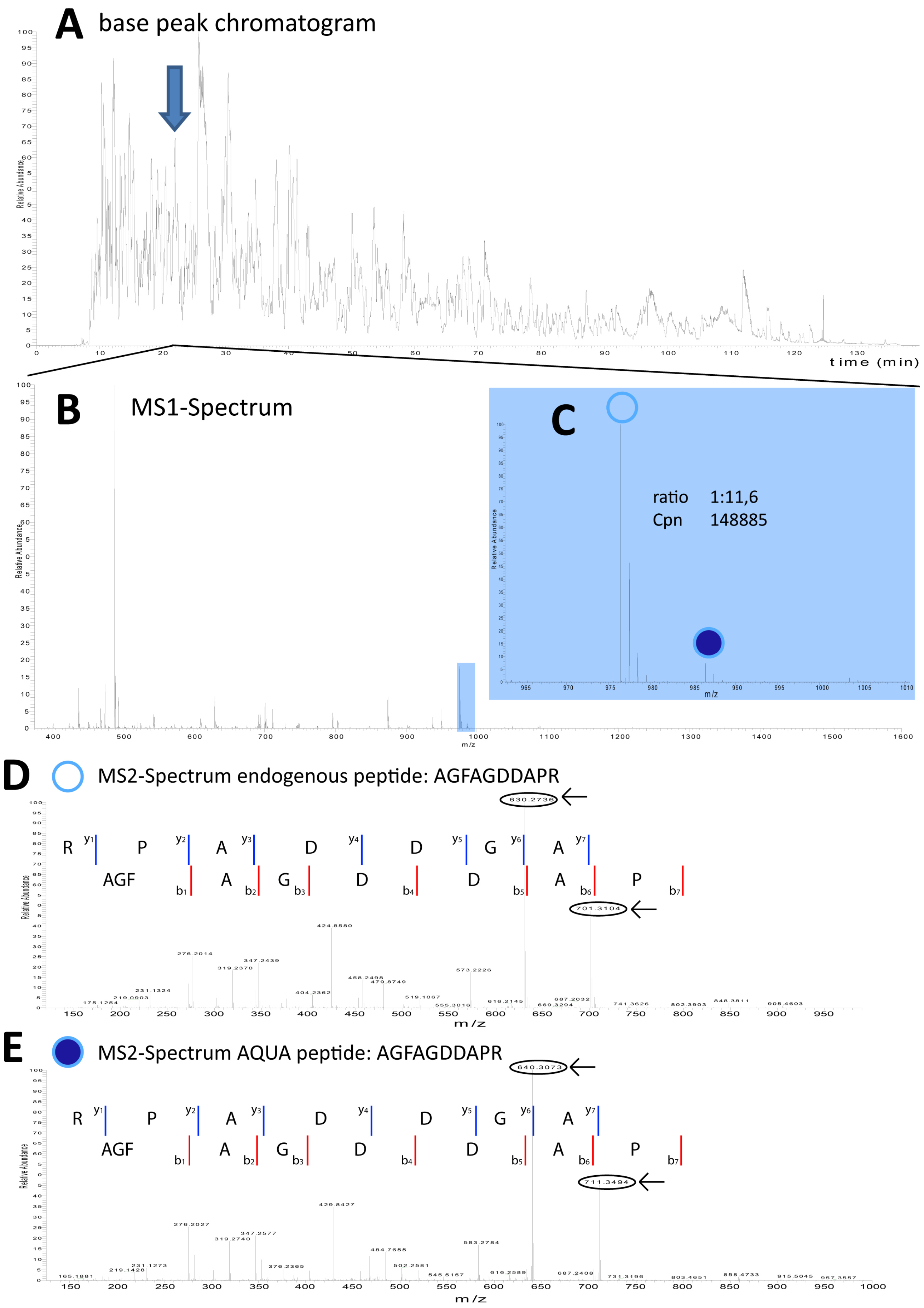
Prior to the LC-MS analysis of tryptic peptides originating from a known amount of spore proteins, a mixture of heavy isotope-labeled reference peptides (AQUA peptides) was added, each in a predetermined precise concentration. The reference peptides were synthesized with heavy 13C and 15N isotope labeled Arginine or Lysine, which results in a mass shift of 10 dalton and 8 dalton, respectively. Due to the same chemical characteristics and ionization properties of the natural peptide and the synthetic heavy peptide both behave identically during the LC-MS analysis, which allows to identify and quantitatively compare the natural and the heavy peptides.

Table2-17 lists the 40 proteins from which AQUA peptides were selected as standards for quantifications. Column 4 shows the total number of peptides detected for each of these reference proteins during MS/MS runs which serves as a first indication of the abundance of these proteins in spores. This number is very high for actin or the fatty acid synthetase FAS1 and very low for the inositole phosphatase INP54 or the kinesin motor Kip2. The sequences listed in column 6 represent the heavy peptides selected for each protein. The purity of these AQUA peptides obtained from Thermo Scientific is 97% and the isotopic enrichment 99% as controlled via MALDI MS and analytical HPLC. Peptide 25 selected for the microtubule-binding protein Bik1 (AFR702W) was not found with tandem mass spectrometry. Also peptides 30 and 37 selected from the endocytotic proteins Pan1 and Abp1, respectively, could not be detected for some unknown reason. The peptide sequence EAEEEDEEEDNAGLK from the SH3 domain-containing protein Bbc1, could not be synthesized with an isotope label. This means for 36 selected proteins direct quantifications based on known amounts of AQUA peptides could be performed. In addition, these primary, AQUA peptide-supported data for protein quantifications for a limited number of proteins were used to establish a regression curve to adjust the quantification of all other proteins in spores as described in the last section.

One example for the different steps in detecting and analyzing the native and the corresponding heavy AQUA peptide is shown for the quantification of actin (Act1) in Figure2-18. The first output is the base peak chromatogram (Figure2-18A) which includes all MS1-Spectra of the 140 minute MS run. Each discovered peptide in the spectra of the chromatogram has a corresponding identification number. The base peak chromatogram originates from the first fraction of the Off-Gel of the spore sample. Since the peptides of the spore sample were first separated into 12 fractions according to their isoelectric points, 12 different runs with matching base peak chromatograms exist. The MS1-Spectrum 2,579 (Figure2-18B) of the chromatogram 02120 (Figure2-18A) represents all ions that were detected at 21.31 minutes after the first sensed ion in this MS run. Ions between 375 and 1,600 dalton are displayed in the MS1-spectrum. Completely separated MS1 peaks are clearly visible in a scale-up of the area at 980 dalton (Figure2-18C). The left peak marked with an open circle has a mass of 975.411 dalton matching the peptide sequence AGFAGDDAPR of the *A. gossypii* protein Act1. Also the distribution of the naturally occurring carbon isotope 13-C within this peptide can be observed by increased shifts of 1 dalton for each 13-C atom. The filled circle in this spectrum marks the heavy homologue of the endogenous peptide with a mass of 985.4493 dalton. This AQUA peptide was synthesized using arginine carrying exclusively heavy nitrogen and carbon isotopes. The peptide AGFAGDDAPR has one C-terminal arginine (R) that leads to an increased mass of the AQUA peptide by 10 daltons, because arginine is composed of 6 carbon and 4 nitrogen atoms and each isotope shifts the peptide by 1 dalton. The exact amount of the natural homologue can now be calculated from the ratio of the heavy and light peptide as outlined in Table2-18 line3. The measured abundances of the AQUA and the natural peptide are 505 and 5856, respectively, which is a ratio of 1:11.6. 20 pmol of the AQUA peptide were added into the complex peptide mixture, which therefore contained 232 pmol of the native peptide which originates from 232 pmol actin isolated from $9x10^8$ spores. By taking this into account each spore carries about 150,000 copies of actin. Table2-18 summarizes these calculations for all proteins for which native and AQUA-peptides were detected as discussed below.

In order to confirm the sequence of the actin reference peptide the MS2 spectrum of the light peptide and the heavy peptide are mapped (Figure2-18D and E). Both spectra show all ions that were detected after peptide fragmentation. The b ions are marked in red and are fragmented from the N-terminus to the C-terminus (see chapter1 and Figure1-4). Trypsin is cutting after every arginine and lysine. So each peptide contains one of these amino acids at the C-terminus. By comparing the b ions of the light and heavy spectrum no differences were detected and should not be detected, because the isotope labeled arginine has to be at the C-terminus. The b1 ion is composed of the amino acids alanine, glycine and phenylalanine, together they have a mass of 275.2 dalton. The matching peak is clearly detected in both MS2-spectra and indicates a charged ion. The second b ion has an additional amino acid alanine. This ion included then 4 amino acids ADFA and has a mass of 346.3 dalton. Also the following b ions were detected; b3 + Gly with a mass of 403 dalton, b4 + Asp and b5 + Asp both have additional 115 dalton. Ion b6 caries another Ala with a total mass of 704 daltons and the last b ion b7 is still quantifiable at 803 daltons. Since peptides are fragmented the full length peptide is not found in the spectra and a shift of the last possible b ion cannot be observed. In contrast to the b ions y ions are fragmented form the C-terminus to the N-terminus. Because of the C-terminal arginine a shift of 10 dalton appears for the heavy peptide for all y ions. An example is presented for the ion y6. The comparison of the heavy and the light ions shows a shift from 630.3 dalton (Figure2-18D) to 640.4 dalton (Figure2-18E), indicated by the ellipse mark around the peak mass and by the arrow. A second example is shown for the next y ion that has the sequence AGDDAPR with a mass of 701 dalton which shifts to 711 dalton for the isotope labeled peptide. The two MS2 spectra demonstrate the good sequence coverage with y and b ions and demonstrate the differences and similarities between the heavy and light peptide. The differences of the y ions are more clearly seen in Figure2-19 which represents an overlay of both the heavy and light MS2 spectra. The fragmentation spectrum of the heavy peptide is colored in red and the y ions that are identified as light are presented in blue. The 10 dalton shift is seen for each y ion. The b ions that carry no heavy isotope in both peptides are marked with black peaks.

The comparison of all other AQUA peptides with their native homologues is summarized in Figure2-20. The abundance for all AQUA peptides that were added with 2 pmol into the peptide sample is indicated as red bars in Figure2-20A. The range of measured abundances is 0.7 to 111 for all synthesized peptides that have the same concentration of 2 pmol. Because of the sequence-dependent ionization abilities peptides with the same mass can show different ionizations, and the peptide with higher ionization efficiency reaches the detector at a higher quantity. With a log ratio average of 0.20 the peptides are relatively close in their abundance distribution. The corresponding native peptides were found in an abundance range from 0.2 to 221 indicated as blue bars. Figure2-20B shows a similar graph for the AQUA peptides spiked with a concentration of 20 pmol and their corresponding native peptides. The abundance of the heavy peptides ranges from 28 to 2,293 and has a recorded log ratio of 0.19. The highest variance from the log abundance mean is 1.38. 16 of the 18 heavy peptides are located in a variance of a 0.5 log abundance compared with the mean. All measured abundances for the heavy and light peptides and their ratios are listed in Table2-18.



A base peak chromatogram

B MS1-Spectrum

C

D MS2-Spectrum endogenous peptide: AGFAGDDAPR

E MS2-Spectrum AQUA peptide: AGFAGDDAPR

Figure2-18: Detection and analysis of the light and heavy reference peptide for Act1. **(A)** The peak based chromatogram of all detected ions in the Off-Gel fraction 1 during the 140 minutes tandem mass spectrometry run. The arrow points to the peak of the MS1-Spectrum shown in **(B)**, in which the heavy and the light peptide are detected. The MS1-Spectrum is presented with a mass range from 400 to 1,600 dalton. The area between 965 and 1,010 dalton, marked in blue, is magnified in **(C)** to show the resolution of the AQUA peptide (filled circle) and the light peptide (open circle) with adjacent natural 13C variants. See result text for further details. **(D)** MS2-Spectrum of the natural peptide K.AGFAGDDAPR.A. **(E)** MS2-Spectrum of the heavy peptide K.AGFAGDDAPR[166.11].A. Both MS2 spectra have a mass range of 150 – 1,000 dalton. Peaks of y-Ions are marked in blue and b-Ions are labeled in red. The equivalent amino acid sequence is written in between the peaks. See result text for further detail.

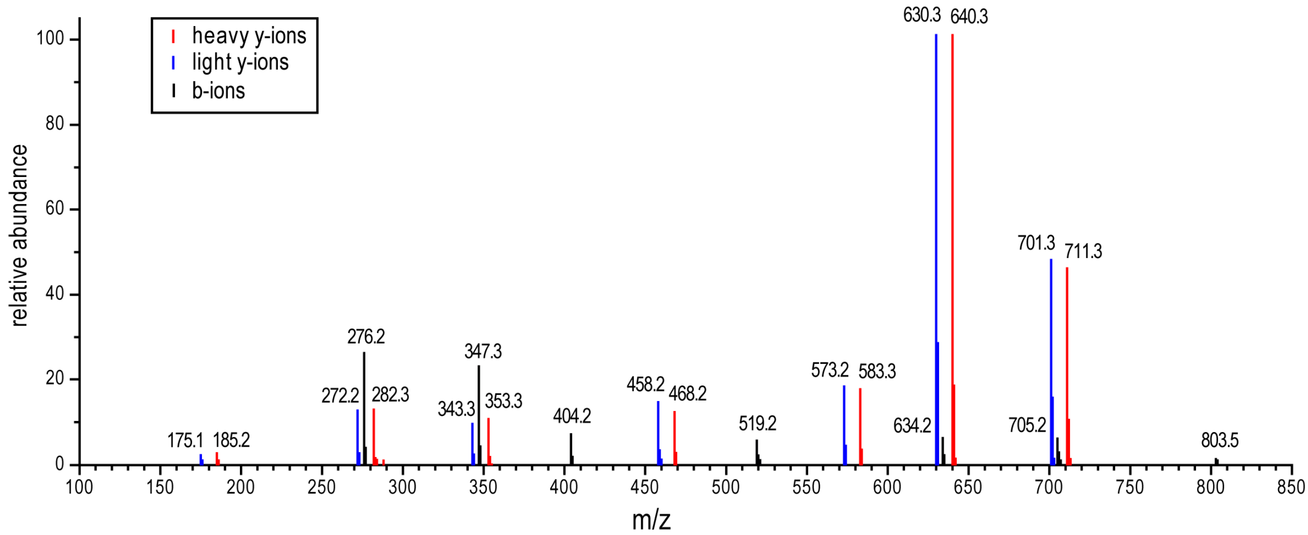


Figure2-19: Overlay of the MS2-Spectra of the heavy and the light Act1 reference peptide. Peaks of y-Ions of the AQUA peptide are colored in red and those of the endogenous peptide in blue. B-Ions are stained in black. The relative abundance (y-axis) is expressed in percent of the highest peak. The mass (x-axis) ranges from 100 to 850 dalton.

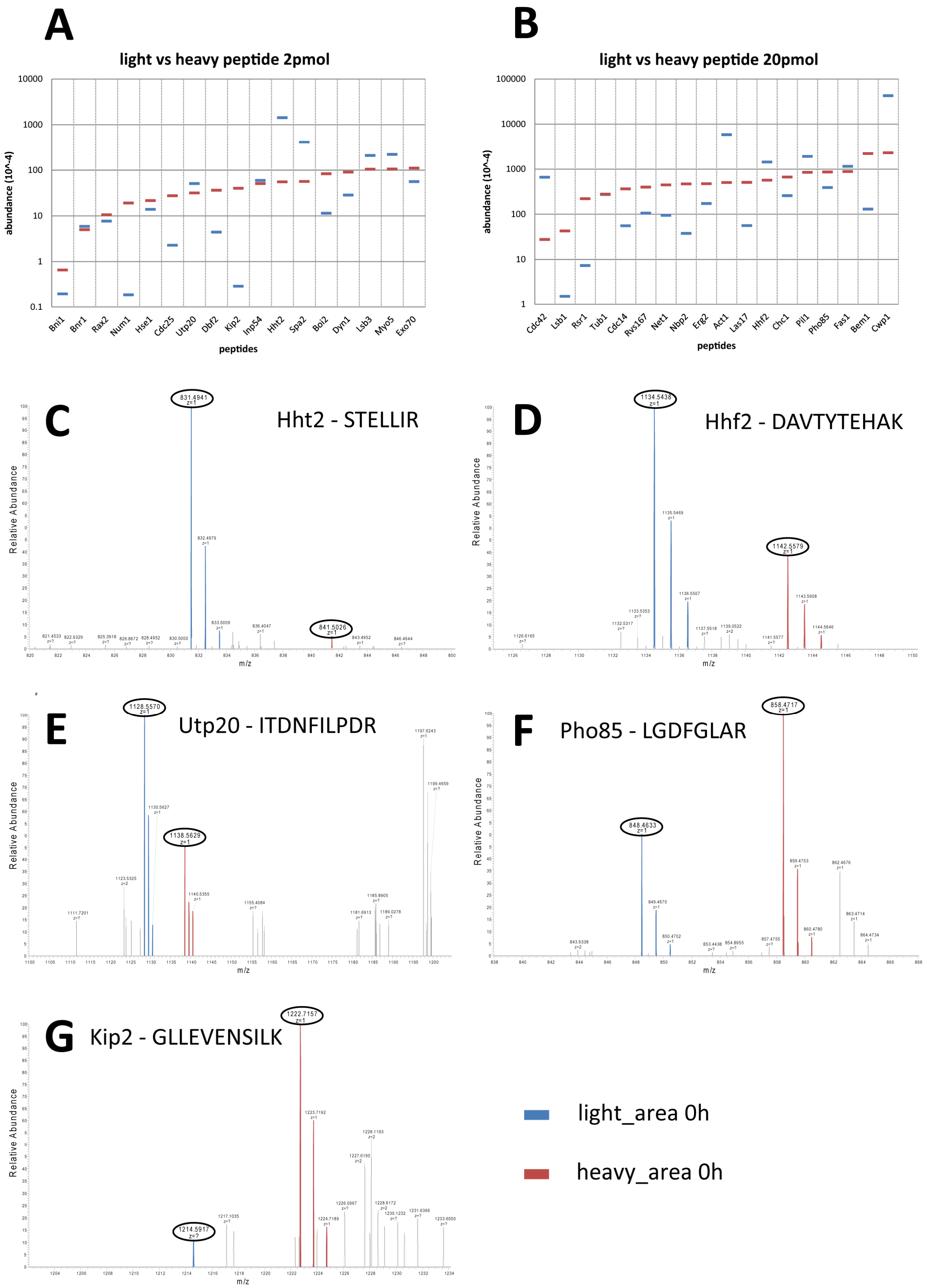

A
light vs heavy peptide 2pmol
abundance (10^-4)
10000
1000
100
10
1
0.1
Bni1
Bnr1
Rax2
Num1
Hse1
Cdc25
Utp20
Dbf2
Kip2
Inp54
Hht2
Spa2
Boi2
Dyn1
Lsb3
Myo5
Exo70
peptides
B
light vs heavy peptide 20pmol
abundance (10^-4)
100000
10000
1000
100
10
1
Cdc42
Lsb1
Rsr1
Tub1
Cdc14
Rvs167
Net1
Nbp2
Erg2
Act1
Las17
Hhf2
Chc1
Pil1
Pho85
Fas1
Bem1
Cwp1
peptides
C
Hht2 - STELLIR
Relative Abundance
m/z
831.4941 z=1
841.5026 z=1
D
Hhf2 - DAVTYTEHAK
Relative Abundance
m/z
1134.5438 z=1
1142.5579 z=1
E
Utp20 - ITDNFILPDR
Relative Abundance
m/z
1128.5570 z=1
1138.5629 z=1
F
Pho85 - LGDFGLAR
Relative Abundance
m/z
848.4633 z=1
858.4717 z=1
G
Kip2 - GLLEVENSILK
Relative Abundance
m/z
1222.7157 z=1
1214.5917 z=?
🟦 light_area 0h
🟥 heavy_area 0h

Figure2-20: Abundances for 35 pairs of heavy and light peptides determined from MS1-spectra. **(A)** Red bars: Abundances of the 17 AQUA peptides added as 2 pmol. Blue bars: Abundances of the corresponding light peptides. **(B)** Red bars: Abundances of the 18 heavy peptides added as 20 pmol. Blue bars: Abundances of the corresponding light peptides. The proteins from which the reference peptides were selected are indicated at the bottom of the graphs. **(C to G)** Examples for sections in MS1-spectra showing heavy and light peptides. All peptide ions are singly-charged, with a mass shift of 1 dalton. Peptide peaks matching endogenous peptides are colored in blue and the isotopic labeled homologues in red. **(C)** Light reference peptide for Hht2 detected at the calculated mass of 830.4861 dalton and the heavy peptide (added as 2 pmol) with a 10 dalton higher mass found in fraction 5 (24.25 min) with a maximal MS1-Spectrum intensity of 1.99x107. The ratio between heavy and light is 25.6. **(D)** Light reference peptide for Hhf2 detected at the calculated mass of 1,133.5353 dalton and the heavy peptide (added as 20pmol) with an 8 dalton higher mass found in fraction 4 (10.48 min) with a maximal MS1-Spectrum intensity of 9.94x106. The ratio between heavy and light is 2.5. **(E)** Light reference peptide for Utp20 detected at the calculated mass of 1,127.5472 dalton and the heavy peptide (added as 2 pmol) with a 10 dalton higher mass found in fraction 7 (19.38 min) with a maximal MS1-Spectrum intensity of 3.22x105. The ratio between heavy and light is 1.6. **(F)** The light reference peptide for Pho85 detected at the calculated mass of 847.4552 dalton and the heavy peptide (added as 20pmol) with a 10 dalton higher mass found in fraction 5 (33.55 min) with a maximal MS1-Spectrum intensity of 1.31x107. The ratio between heavy and light is 0.45. **(G)** Light reference peptide for Kip2 detected at the calculated mass of 1,213.705 dalton and the heavy peptide (added as 2 pmol) with an 8 dalton higher mass found in fraction 2 (63.69 min) with a maximal MS1-Spectrum intensity of 1.25x106. Ratio between heavy and light is 0.0071.

Five additional examples of the MS1 spectra for pairs of heavy and light peptides are shown in Figure2-20C-F. The mass of the light and heavy peptides are marked by an ellipse. The peaks for the light area are colored in blue and the equivalent heavy peptide peaks are colored in red. The MS1 spectrum of the light and heavy peptide STELLIR of the histone Hht2, documented in Figure2-20C, shows the highest ratio difference between an AQUA peptide and its light homologue. The peak maximum of the light peptide is the most intense ion in this part of the zoomed-in spectrum between 820 and 850 Dalton, and was scaled to 100%. The peak maximum of the heavy peptide, shifted by 10 dalton, was only 4% of the most abundant ion detected in this part of the scan. The AQUA peptide was spiked into the complex peptide sample with a concentration of 2pmol. The localization in the sequence of the Hht1 peptide is in the central part of the protein and marked in red in Figure2-12 (Chapter2-B). The AQUA peptide DAVTYTEHAK (Figure2-13) chosen for the quantification of histone Hhf2 is similarly abundant as the light peptide (Figure2-20D). This is not surprising because it was injected with an amount of 20 pmol. In this case the mass of the heavy peptide DAVTYTEHAK is shifted only by 8 dalton from 1,134.54 to 1,142.56 dalton compared to the natural peptide, because the C-terminal lysine carries 6 heavy carbon and 2 heavy nitrogen atoms. The comparison of Figure2-20C and D shows the differences between the two concentrations of AQUA peptides. Both light peptides of Hht2 and Hhf2 have the same peak area, in other words were found with very similar abundance as documented in Table2-18 and as expected for these two histones which should be present in equimolar amounts. The tenfold difference of the corresponding heavy peptide abundances can be fully accounted for by the tenfold difference in the amount spiked into the peptide mixture.

Figure2-20E and F illustrate the quantification of the proteins Utp20, involved in 18SrRNA biogenesis, and Pho85, a cyclin-dependent kinase. The isotopic labeled reference peptide HGNQFEQLR for Utp20 has only a twofold lower abundance then its native homologue (Figure2-20E). The mass difference of the heavy and the light peptide is 10 dalton (1,128.56 to 1,138.56 dalton). The Pho85 reference peptide has a twofold higher abundance compared to its light homologue (Figure2-20F). The heavy peptide LGDFGLAR [166.11] changes the mass by 10 dalton (848.46 to 858.47 dalton) as expected for the C-terminal arginine. An extreme example for a much higher abundance of the heavy peptide compared to the light homologue was found for the kinesin motor Kip2 (Figure2-20G). The heavy peptide GLLEVENSILK is about 150 times more abundant than the native homologue which allows only a good estimate of the Kip2 copy number in spores. The very low abundance of the light peptide prevents the detection of the natural isotopic variants seen for all other peptides discussed in this section.

Table2-18 summarizes all data to calculate copy numbers per spore for the 35 proteins. The light peak intensities are listed in column 3, the heavy peak intensities in column 4, their ratios in column 5, the concentration (pmol per 2 mg spore protein) of the native peptides in column 7. Taking the spore quantification (Figure2-17) into account, with 2 mg proteins extractable from 9.4×10^8 spores (column 8) we transformed the result into protein concentration per spore (column 9). For conversion the concentrations into copy numbers per spore, first the pmol were transformed into mol by multiplying with 10^{-12} and then multiplied with the Avogadro number (6.022×10^{23}). The final result is copies per spore (column 10) for each protein with an isotope-labeled reference peptide.

The biological validation of the copy numbers for the 35 proteins reveals some interesting details. Each of the core histones Hht2 (H3) and Hhf2 (H4) was found with approximately 32,500 copies per spore. These core histones form together with the histones H2B and H2A the nucleosomes an octamer protein complex (Schafer et al., 2005). Each nucleosome is wrapped by about 150 base pairs DNA, and with the additional linker histone Hho1 that can generate space between 10 and 90 bp (Richmond and Davey, 2003), there are at most 240 bp per nucleosome. With about 9 megabases the *A. gossypii* genome is the smallest genome of all characterized free-living eukaryotes (Dietrich et al., 2004). The total genome size of 9×10^6 bp divided by the maximal nucleosome distance of 240 bp results in a nucleosomes number of 38,000 per nucleus. Each nucleosome contains two copies of the histone H3 and H4 heterodimers. Therefore, the calculated copy number per nucleus of 32,500 for Hht2 and Hhf2 is two times lower than the expected number if the DNA in spores is packaged as nucleosomes including promoters and other inter ORF regions. This estimation is based on vegetative cells, but as shown by Hegemann and Wilmen the nuclease cleavage of chromatin in *S. cerevisiae* spores also shows the typical nucleosome digestion pattern observed for chromatin in vegetative cells (Wilmen et al. 1996). The calculated number of nucleosomes is 16,000 based on the AQUA peptides, which is relatively close to the biological estimation. With a closer look at the histone proteins and their copy number calculations via the regression line (see below) using the All-method, the Hht2 protein has a very different value compared to the AQUA peptide calculation and also relative to the Hhf2 copy numbers. Only around 2,000 copies were calculated with the All-method compared to the expected 32,000 copies.

Can this be explained with the high modification level of histone H3? By just comparing the AQUA peptides with the endogenous peptide the only peptide that is localized in the central part of the histone sequence (Figure2-13) that is very highly modified is also critical for the concentration determination. Furthermore modification like acetylation and methylation are primarily linked to lysines, blocking cleavage by trypsin which leads to partially cleaved di- or tri-peptides (see Chapter2-B). With the loss of the recognition of the cleavage site the ions can become too large to become detected by LTQ Orbitrap. Due to the sequence of the *A. gossypii* H3 histone, peptides in the detection range are rare precluding a realistic estimate of its abundance. For abundance optimization of histones see Table2-14 in part B of chapter2.

The Num1 protein, with a low but exactly defined protein copy number per spore of 12, interacts with the motor protein dynein (Tang et al., 2009) and is important for the dynamics of the microtubule cytoskeleton. In *A. gossypii* Num1 mediates the interaction of the microtubules with the cell cortex via dynein, which can result into pulling forces (Grava and Philippsen, 2010). In spores neither active microtubule pulling nor polymerization and depolymerization is expected to take place. This low copy count indicates no need for dynamic microtubules and their function in nuclear migration (Alberti-Segui et al., 2001; Lang et al., 2010) because the nucleus is in a stable and fixed position in the spore. Like the Num1 also the Kip2 protein interacts with microtubules. Kip2 is a kinesin motor and transports the plus-end binding protein Bik1 to the plus ends of cytoplasmic microtubules and helps to control growth of the microtubules (Carvalho et al., 2004; Grava and Philippsen, 2010). The low Kip2 copy number of 9 per spore is in the same range like Num1 and confirms the quality of the data and suggests a low number of microtubules. Interestingly, the copy number of Dyn1, the main component of the dynein motor, is 40 times higher in spores than Num1 and Kip2. Dynein may be stored in a higher copy number because it is also used for nuclear microtubules during germination because mitoses already occur in small germ bubbles (Chapter3).

Cwp1 is the 10th highest abundant protein in the spore proteome (Table2-15). Using the direct quantification with an AQUA peptide a copy number of 240,000 per spore was calculated (Table2-18). Cwp1 is involved in the cell wall organization and codes for a mannoprotein which is linked to beta-1,3-glucan and beta-1,6-glucan polymers via a phosphodiester bond (Kapteyn et al., 1999). The genome of *A. gossypii* carries 7 homologues of the *S. cerevisiae* Cwp1 protein originating from tandem and non-tandem gene amplifications (Dietrich et al., 2004) implying modifications in the cell wall organization of both yeasts (Schmitz and Philippsen, 2011). The systematic name of Cwp1 (d) is ABR028C. The protein could play an important role in building the spore cell wall or it is an essential protein needed for the germination of spores when new cell walls must be synthesized as fast as possible. The availability of this Cwp1 variant would be necessary to guarantee the outgrowth of the spore without leakage of the germ bubble. It should be noted that the mRNA for his Cwp1 variant dramatically decreases during the development from spores to hyphae (Riccarda Richatsch, PhD-thesis 2008).

Table2-18: Copies per spore of selected proteins for which AQUA peptides were synthesized. From the measured peak intensities in MS1-Spectra (columns 3 and 4) and the concentration of the added AQUA peptides the protein concentrations were calculated (column 7). Based on 2.5pg protein per spore the copy numbers per spore could be calculated (bold numbers column 10). Additionally, the data in column 11-15 are used for establishing a regression curve to adjust the copy numbers of all detected spore protein (figure2-21).

		relative abundance												
accession	common name	light area	heavy area	ratio from area	heavy peptide amount in[pm]	concentration in pmol	spore count	protein concentration in pmol per nucleus	**copies per nucleus**	protein abundance / possible peptides [10^-4]	average protein abundacne of top 3 peptides [10^-4]	log copies per nucleus	log protein abundance / possible peptides	log protein abundacne top 3 peptides
AGL093W	Cdc42	661.18	27.69	23.8779	20	477.5587	937798901	5.09234E-07	**306660**	873	2946	*5.487*	*6.941*	7.469
ABR028C	Cwp1	42860.59	2292.76	18.6939	20	373.8784	937798901	3.98677E-07	**240083**	18324	102154	5.380	8.263	9.009
ABR222W	Act1	5856.15	505.15	11.5928	20	231.8569	937798901	2.47235E-07	**148885**	15591	107018	5.173	8.193	9.029
ADL202C	Hht2	1424.00	55.60	25.6115	2	51.2230	937798901	5.46205E-08	**32892**	97	153	*4.517*	*5.988*	*6.184*
ADL201W	Hhf2	1430.00	568.20	2.5167	20	50.3344	937798901	5.36729E-08	**32322**	1623	2697	*4.509*	*7.210*	*7.431*
AEL329W	Pil1	1911.50	847.00	2.2568	20	45.1360	937798901	4.81298E-08	**28984**	4037	30170	4.462	7.606	8.480
AER085C	Fas1	1145.72	881.32	1.3000	20	26.0003	937798901	2.77248E-08	**16696**	1369	18683	4.223	7.136	8.271
AFR441C	Tub1	273.65	279.37	0.9795	20	19.5907	937798901	2.08901E-08	**12580**	818	3069	4.100	6.913	7.487
ADL022C	Spa2	414.20	56.48	7.3336	2	14.6671	937798901	1.564E-08	**9418**	3.8	528	*3.974*	*4.583*	6.723
AGL242C	Pho85	389.68	861.90	0.4521	20	9.0423	937798901	9.6421E-09	**5806**	122	1059	3.764	6.088	7.025
AER359W	Chc1	258.80	668.00	0.3874	20	7.7485	937798901	8.26244E-09	**4976**	352	4703	3.697	6.546	7.672
AFL181C	Erg2	173.39	473.84	0.3659	20	7.3185	937798901	7.80392E-09	**4700**	240	1079	3.672	6.381	7.033
AFR140C	Rvs167	105.89	399.68	0.2649	20	5.2987	937798901	5.65015E-09	**3403**	156	1311	3.532	6.194	7.118
AAL181C	Net1	94.66	446.51	0.2120	20	4.2401	937798901	4.52132E-09	**2723**	67	2143	3.435	5.824	7.331
AEL306C	Myo5	220.94	106.02	2.0839	2	4.1679	937798901	4.44434E-09	**2676**	80	9352	3.428	5.901	7.971
AEL017W	Lsb3	209.89	105.20	1.9952	2	3.9904	937798901	4.25504E-09	**2562**	210	1706	3.409	6.322	7.232
AEL006W	Utp20	51.00	31.80	1.6038	2	3.2075	937798901	3.42029E-09	**2060**	0.72	90	*3.314*	*3.857*	5.956
AEL025W	Cdc14	55.70	365.40	0.1524	20	3.0487	937798901	3.25092E-09	1958	66	837	3.292	5.822	6.923
AGL364C	Bnr1	5.90	5.01	1.1776	2	2.3553	937798901	2.51151E-09	**1512**	0.0045	0.17	*3.180*	*1.657*	*3.237*
AEL090C	Inp54	59.60	50.80	1.1732	2	2.3465	937798901	2.50209E-09	**1507**	33	356	3.178	5.512	6.551
AGR285W	Las17	56.37	506.90	0.1112	20	2.2241	937798901	2.37163E-09	**1428**	19	179	3.155	5.285	6.253
AGL169C	Nbp2	37.68	470.00	0.0802	20	1.6034	937798901	1.70975E-09	**1030**	121	364	3.013	6.082	6.561
AGR095W	Rax2	7.70	10.60	0.7264	2	1.4528	937798901	1.54919E-09	**933**	12	243	2.970	5.071	6.386
ABR008C	Hse1	13.89	21.60	0.6431	2	1.2861	937798901	1.37141E-09	**826**	230	2254	2.917	6.362	7.353
AEL241W	Bem1	130.13	2199.59	0.0592	20	1.1832	937798901	1.2617E-09	**760**	165	928	2.881	6.217	6.967
AFR100W	Exo70	56.20	111.00	0.5063	2	1.0126	937798901	1.07978E-09	**650**	37	378	2.813	5.573	6.578
AFR320W	Lsb1	1.50	42.87	0.0350	20	0.6993	937798901	7.45707E-10	**449**	34	158	2.652	5.537	6.200
AFR464W	Rsr1	7.35	220.49	0.0333	20	0.6666	937798901	7.10781E-10	**428**	18	175	2.631	5.245	6.244
ACR258W	Dyn1	28.53	90.60	0.3149	2	0.6298	937798901	6.71574E-10	**404**	6.4	2986	2.607	4.806	7.475
AFR669W	Bni1	0.19	0.66	0.2951	2	0.5902	937798901	6.29393E-10	**379**	5.9	232	2.579	4.768	6.365
AGL293C	Boi2	11.45	83.33	0.1375	2	0.2749	937798901	2.93155E-10	**177**	14	364	2.247	5.161	6.561
ADR033W	Dbf2	4.42	36.40	0.1214	2	0.2429	937798901	2.58965E-10	**156**	7.4	94	2.193	4.869	5.972
ADL038W	Cdc25	2.28	27.58	0.0827	2	0.1653	937798901	1.76303E-10	**106**	14	884	2.026	5.144	6.946
AGR043W	Num1	0.18	19.10	0.0097	2	0.0193	937798901	2.06119E-11	**12**	4.9	834	1.094	4.690	5.954
ACR145W	Kip2	0.28	40.20	0.0071	2	0.0141	937798901	1.50665E-11	**9**	0.012	0.40	*0.958*	*2.077*	*3.602*

Generation and use of an AQUA peptide-based regression line

The 35 protein copy numbers exactly determined from the MS intensities of AQUA peptides and their native homologues cover over four orders of magnitude. These data should therefore be well suited to establish a regression line to directly translate all determined MS intensities (or protein abundances) into protein copy numbers per spore. First, protein copy numbers were plotted against the protein abundances estimated by two methods from MS intensities (Figure2-21 A and B). These scatter plots were drawn in a log scale and the logarithmic values were taken from Table2-18 columns 13 to 15. The two methods suitable for estimation of protein abundances were extensively described in Chapter 1. In the iBAQ approach, also called All-method in this PhD-thesis, the MS intensities of all detected peptide ions of a protein were summed and divided by the number of possible tryptic peptides suitable for detection by LC-MS (y-axis in panel A). In the T3PQ approach, also called Top3-method in this PhD-thesis, the average MS intensity of the 3 most intense peptides for each protein was used to define the abundance (y-axis in panel B). The linear fit in panel A was deduced from 29 of the 35 proteins, since some proteins identified with low confidence or a standard deviation of more than 2-fold were not considered. For panel B the linear fit was generated under the same condition and parameters like in the first method using 28 of the 35 reference proteins. The slope of the regression line for the iBAQ method is 0.97 and for the T3PQ method 0.78, which results in a flatter curve. The quality of these regression lines can be estimated by the R-Square (R^2) coefficient. The R^2 for the iBAQ values was 0.844 and therefore higher compared to 0.716 calculated for the T3PQ values indicating that the iBAQ values match better with the copy numbers determined with AQUA peptides.

To further compare the results of the iBAQ and T3PQ approach, we assessed the expected error rates using a bootstrap method (Efron and Tibshirani, 1991) for both data sets (Figure2-21C and D). This approach provides good error estimation and is used to resample rows of a data set. We calculated the regression line for the observed data where each data point was randomly left out from the original dataset, and a regression line was defined for all possibilities. Sampling all variances of these population results in an approximate distribution. The regression lines for each of these estimations are drawn in different colors in panels C and D. The error calculation for both bootstrapped data sets is shown in Figure2-21E and F. Compared with the discussed R^2 coefficients, the error calculation via bootstrapping confirmed the superior quality of the iBAQ method. More specifically, the average fold error of the iBAQ method was only 77.9% compared to 162.2% for the T3PQ strategy.

We translated the relative abundances of the reference proteins obtained with the All-method and Top3-method into protein copy numbers using the formula of the regression lines established in panels A and B, respectively, of Figure2-21. Then we compared the calculations via the regression line with the copy numbers obtained via the ratios between heavy and light peptides. The results are summarized in Table2-19. This table lists in columns 3 to 5 the absolute protein copy numbers and the estimated protein copy numbers per spore for the iBAQ and T3PQ method, and in columns

6 and 7 the ratios of the AQUA peptide-based copy numbers and those derived from the IBAQ or T3PQ method, respectively. The mean shown at the bottom of Table2-19 for both sets of ratios was calculated without outliers that are indicated in light gray and italics. As evident, the iBAQ method shows ratios close to 1 and is therefore well suited to translate MS intensities of peptides into protein abundances. This can probably be ascribed to the fact that in contrast to the T3PQ method the iBAQ method considers all detected peptides for protein quantification, which increases the number of data points and reduces analytical variability.

Nonetheless, as shown in Table 2-19, for some proteins, the estimated protein copy numbers using the iBAQ method were several fold different to the precisely determined protein copy numbers via AQUA peptides. For instance, using AQUA peptides histones H3 and H4 (Hht2 and Hhf2) have almost identical protein copy numbers but their copy numbers determined with the iBAQ method differ by a factor of 18 (Table2-19, lane4). This can be caused by post-translational protein modifications or low ionization efficiencies of the few generated peptides. In total there are only 7 cases where the calculated copies per spore, using the iBAQ method, differ by more than a factor of 10 compared with the protein copy numbers that were calculated via the AQUA peptides. For 15 of the 35 proteins the copy numbers per spore determined by both methods differ by less than a factor of 2. For example, Exo70 has 650 proteins copies per spore calculated with the AQUA peptide and 643 copies determined via the regression line. Additionally, 12 copies were found for the Num1 protein and the evaluation with the iBAQ method revealed 14 proteins per spore confirming the quality of the method. In this case an unusually high number of 46 peptides were detected for Num1 with tandem mass spectrometry.

Copies per spore for all detected protein

We applied the regression line of panel A of Figure2-21 to translate all abundances based on the All-method into copy numbers per spore. Similarly, we applied the regression line of panel B of this Figure to translate all abundances based on the Top3-method. Table2-20 (Appendix 5) lists all protein copy numbers per spore obtained with and without the use of a regression line. In column 3 of this table the values of the regression line calculation with the iBAQ method are shown. Column 4 contains the copies per nucleus calculated with the per cent protein abundance based on the All-method. In column 5 the values of the regression line calculation with the T3PQ-method are shown. Column 6 contains the copies per nucleus calculated with the per cent protein abundance based on the Top3-method. Some proteins are present in over one million copies per spore. In total there are approximately 40 million proteins in one spore. Among the four procedures the copy numbers differ in some cases dramatically, but for most proteins similar copy numbers were determined. For some proteins with a very low coverage of identified peptides copy numbers below one were calculated. But this is the case for only about 100 proteins of the total 3,857 detected proteins in spores.

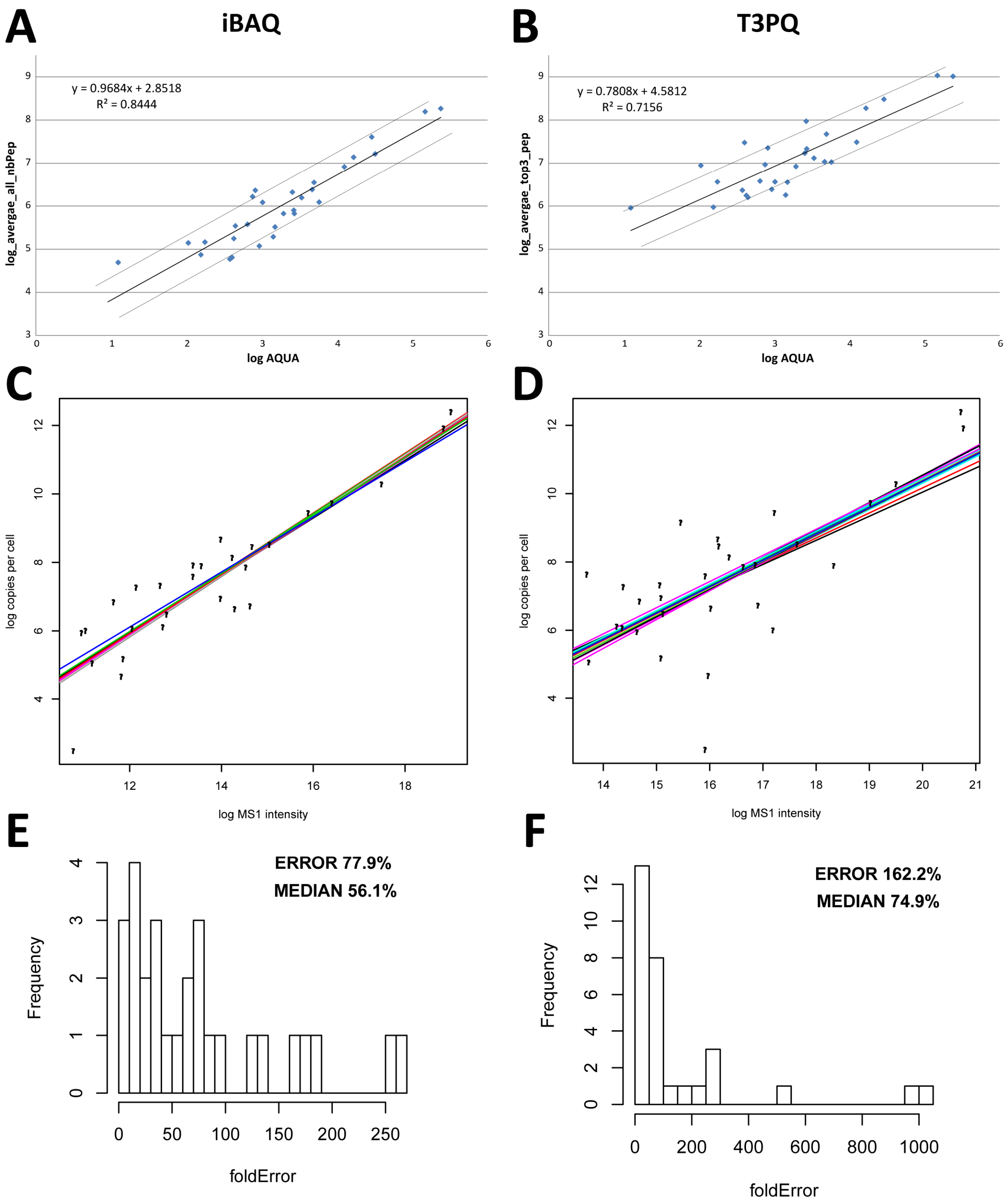


Figure2-21: Generation of regression lines to translate relative protein abundances into protein copies per spore. **(A)** Plotting of the logarithm of relative abundances (y-axis) obtained with the All-method against the logarithm of the copy number accurately determined via the heavy – light peptide ratio (x-axis) of 29 reference proteins. The formula for the linear fit line is shown with errors of +/- 0.07999 for 0.96838 and +/- 0.27212 for 2.85179. R² is 0.8444 and P smaller then 0.0001. **(B)** Plotting of the logarithm of relative abundances (y-axis) obtained with the Top3-method against the logarithm of the copy number accurately determined via the heavy – light peptide ratio (x-axis) of 28 reference proteins. The formula for the linear fit line is also shown with errors of +/- 0.09654 for 0.78076 and +/- 0.32394 for 4.58118. R² is 0.7156 and P smaller then 0.0001. Thin lines in both panels show the borders of plus and minus 0.5 alterations. **(C)** Bootstrap analysis of the data used in A. **(D)** Bootstrap analysis of the data used in B. **(E)** Error calculation of the iBAQ bootstrapping. The average error is 77.9% and the median 56.1%. **(F)** Error calculation of the T3PQ bootstrapping. The average error is 162.2% and the median 74.9%.

Discussion

To quantify absolute concentrations of specific proteins in complex protein mixtures we optimized LC MS/MS run conditions to gain better access to absolute protein copy numbers. We used isotope (13C/15N) labeled internal peptides to quantify the endogenous protein. Furthermore, the isotope dilution technique was used to compute a peptide standard that was used to calculate a protein concentration for all identified proteins. The protein copies per nucleus help us to understand the importance of protein and the protein complexes that a spore needs to form, to survive in the absence of nutrients, and start to germinate in the presence of nutrients.

AQUA peptide standards and native peptide characterization

Using the proteome overview obtained at the start of this thesis with proteins isolated from a mixture of young mycelium (9h incubation) and well advanced mycelium (18h incubation) we were able to select 40 proteins of interest, which had given clearly identifiable MS1 spectra of tryptic peptides. The synthezised heavy isotope-labeled peptides in a given concentration are used to compare the abundance with the native peptides. Both have identical physical and chemical characteristics and due to the small mass shift they form two easily identifiable peaks in the MS1 spectra, which are eluted in the same chromatogram fraction (Figure2-19 and Figure2-20). The comparison of the peak intensities of the heavy peptide and the native peptide, presented in Table2-18 is the best way to quantify the exact amount of an individual protein. For 35 of 40 quantified proteins we determined their copy numbers per spore with two independent methods presented in Table2-19. The data are reinforcing the quality of the methods as shown for example with the histones H3 and H4 which were found in very similar copy numbers. Due to the function of these histones in nucleosomes (Schafer et al., 2005) the protein copy numbers had to be very similar. But there are exceptions in the comparison of the native peptide with the AQUA peptide, as shown for the selected peptide from the protein Cdc42, where we detected a too high concentration. Cdc42 is a small GTPase and is involved in cell polarity (Johnson, 1999). In *A. gossypii* the Cdc42 is essential for establishment of cell polarity (Knechtle et al., 2003). To compare the Cdc42 with other protein localized to the zone of polar growth in *A. gossypii*, we have chosen the SH3 proteins Bem1 and the Boi1. Both are localized to the growing tip of the filamentous fungus *A. gossypii*. Bem1 is a scaffold protein and interacts directly with Cdc42 and with Boi1 (Madden and Snyder, 1998). Cdc42 with a copy number of 300,000 proteins is present at much higher copies compared to Bem1 with 760 or Boi1 with 177 copies per spore. The large differences in the protein copy numbers could be caused by the ionization abilities of the natural peptide or the AQUA peptide was spotted into the complex peptide sample at a concentration of 2 pmol instead of 20pmol. Both could lead to extremely different values for one of the measurements.

Protein copy numbers estimations for all proteins

To estimate protein copies per spore for the whole proteome, we used the heavy peptide standard to compute a regression line, and calculated accordingly the copy numbers from the protein abundances. The protein abundance was based on the All-method or the Top3-method. Date obtained from both methods were refined with a regression line specific for either method. Compared to the T3PQ method, the IBAQ method showed much better linear correlations visible by the higher R^2-values and lower fold errors as determined by bootstraping making it the method of choice for this dataset. Additionally, this could be verified with the median and the mean of the ratio from the direct AQUA peptide determination and the calculation via the linear fit (Table2-19). It is worth mentioning that many protein copy numbers differ only by a factor of two when determined by the iBAQ-method and the AQUA peptide driven method. In contrast the T3PQ method results in a much higher variation of the copy numbers. The advantage of the T3PQ method is that only well ionizable and good flying peptides are used, but partial digestions of the protein and post-translational modifications can result in a much decreased abundance of the MS1 peak. To take all peptides and their accumulated abundances into account is desirable but cannot be achieved for a substantial number of proteins.

Alternatively, we used the approach to calculate protein copies per spore via the percentage of protein abundance in the total protein mixture (Table2-15 and Table2-20). Overall the relative abundances are in the same order compared with the copy numbers determined via a regression line. The numbers for one method may be overall larger than numbers from another method because the picogram protein in one spore, taken as 100%, was experimentally determined by two approaches. Starting with the spore titer we estimated 2 pg protein per spore (Figure2-17) which was used for calculating protein copy numbers via AQUA peptides. Starting with the average spore volume we estimated 2.5 pg to 2.8 pg protein per spore (Figure2-15) which was used to calculate protein copy numbers via relative protein abundances (Table2-2). This latter method may include a relatively high error, but results in a similar protein concentration per single spore.

Table2-19: Comparison of the direct quantification of protein copy numbers using AQUA peptides with two other methods lacking internal concentration standards but using the regression curve established in Figure2-21). *1) All detected peptides were taken into account and the sum of their abundances was divided by the number of all detectable peptides for a protein. *2) The average of the 3 peptides with the highest abundance was used. *3) Mean of the methods with outliers indicated by light gray numbers.

accession	common name	cpn AQUA peptide	cpn all peptides *[1]	cpn top3 peptides *[2]	ratio AQUA / all peptides2	ratio AQUA / top3 peptides
AGL093W	Cdc42	306660	19728	3554	16	86
ABR028C	Cwp1	240083	395951	230553	0.61	1.04
ABR222W	Act1	148885	359229	253191	0.41	0.59
ADL202C	Hht2	32892	1733	113	19	*292*
ADL201W	Hhf2	32322	31677	4464	1.02	7.24
AEL329W	Pil1	28984	96055	77006	0.30	0.38
AER085C	Fas1	16696	28969	40738	0.58	0.41
AFR441C	Tub1	12580	15713	4475	0.80	2.81
ADL022C	Spa2	9418	17	23	*542*	*404*
AGL242C	Pho85	5806	1999	734	2.90	7.91
AER359W	Chc1	4976	6555	6410	0.76	0.78
AFL181C	Erg2	4700	4110	546	1.14	8.60
AFR140C	Rvs167	3403	2834	773	1.20	4.40
AAL181C	Net1	2723	1154	1712	2.36	1.59
AEL306C	Myo5	2676	1320	1033	2.03	2.59
AEL017W	Lsb3	2562	3760	1622	0.68	1.58
AEL006W	Utp20	2060	11	9	*187*	*220*
AEL025W	Cdc14	1958	1179	490	1.66	4.00
AGL364C	Bnr1	1512	0.06	0.02	*25877*	*79593*
AEL090C	Inp54	1507	392	117	3.84	13
AGR285W	Las17	1428	338	82	4.22	17
AGL169C	Nbp2	1030	2168	342	0.48	3.01
AGR095W	Rax2	933	195	84	4.78	11
ABR008C	Hse1	826	4218	2862	0.20	0.29
AEL241W	Bem1	760	2598	1122	0.29	0.68
AFR100W	Exo70	650	643	294	1.01	2.21
AFR320W	Lsb1	449	557	70	0.81	6.45
AFR464W	Rsr1	428	296	64	1.45	6.73
ACR258W	Dyn1	404	92	165	4.38	2.44
AFR669W	Bni1	379	24	21	16	18
AGL293C	Boi2	177	231	154	0.77	1.15
ADR033W	Dbf2	156	121	60	1.29	2.60
ADL038W	Cdc25	106	238	455	0.44	0.23
AGR043W	Num1	12	14	26	0.88	0.47
ACR145W	Kip2	9	0.2	0.1	57	*162*
				mean *[3]	4.63	12.19
				median	1.20	3.01

Chapter 3: Proteome of *A. gossypii* germ bubbles

Chapter 3: Proteome of *A. gossypii* germ bubbles

Introduction

During germination, biological processes are initiated which result in the emergence of isotropic growth in the center of the needle-shaped *A. gossypii* spores. When spores are inoculated in full medium they stay morphologically unchanged during the first two to three hours (Figure3-1A). After three to four hours a small bubble forms in the middle of spores (Figure3-1B). These germ bubbles continue to grow isotropically for a few hours whereas the two needle-shaped cones do not increase in size and stay attached to the bubbles (Figurte3-1C). When germ bubbles reach a critical size after about 6 to 8 hours they switch to the typical constitutive polar growth of hyphae (see Chapter4). Since we wanted to study the proteome of isotropically growing germ bubbles, spores were incubated for only five hours, a time prior to the switch to polar growth. The start of germination in a spore population is an asynchronous process. Therefore, germination cultures consist of a mixture of spores and of germ bubbles with different sizes as concluded from the observed cell morphologies. Some spore-like cells may be non-viable others may start germination at a later time. Spores which do not germinate after nine to eleven hours are counted as non-viable.

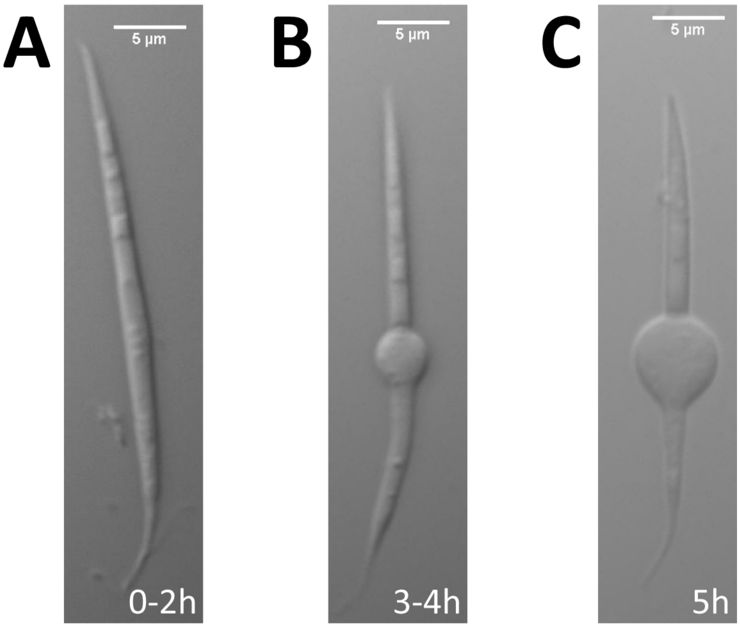


Figure3-1: Germination of *A. gossypii* spores to germ bubbles. **(A)** Needle-shaped spore. **(B)** Germling with an emerging germ bubble. **(C)** Germling with a large germ bubble 5 hours after spore inoculation. Scale bar 5 µm.

The five hours cultures investigated here was a mixture of 35% germ bubbles and 65% spores (Figure3-2). The first sections describe the analysis of the proteome in this mixture of cells which should already reveal increased abundances of several groups of proteins important for growth, and most likely also decreased abundances of proteins only needed for sporulation or spore survival. We employed methods for estimating the relative abundances of proteins but also for absolute quantifications using heavy isotope-labeled reference peptides as described in the previous chapter. For better comparison with the spore proteome we normalized the protein abundances as copy number per nucleus because spores carry one nucleus and germ bubbles can carry several nuclei.

In the last two sections two attempts are presented to extract a more germ bubble-specific proteome from the proteome data representing the mixed cell population. Basically, the copy numbers of proteins in spore-like cells present in the cell mixture were estimated and subtracted from the in most cases higher copy numbers of proteins found in the mixture of cells. This was possible because the proteome of spores was characterized at high resolution in the previous Chapter and because the protein content of spores is known. Thus, it should now be possible to identify germ bubble proteins important for isotropic growth and of proteins necessary for spore germination from proteome data adjusted for germ bubbles.

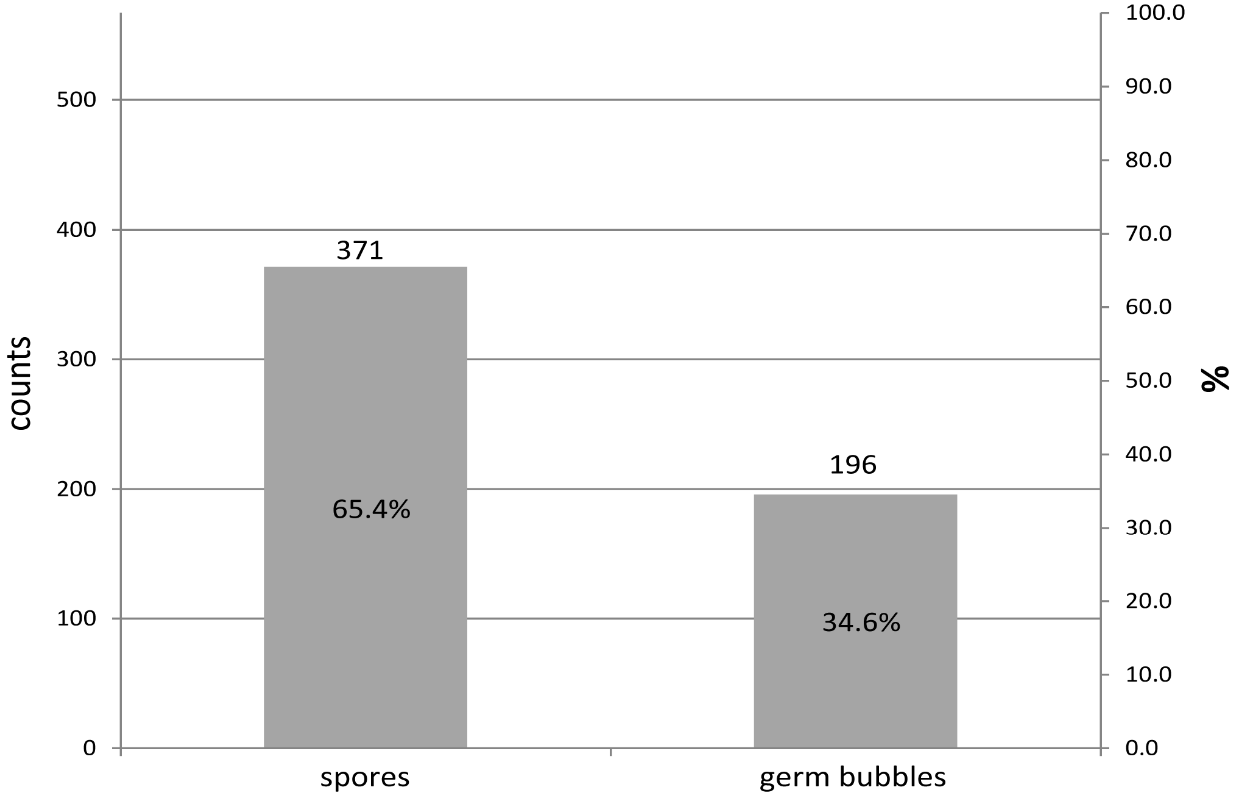


Figure3-2: Composition of cell types based on morphology after 5 hours after spore inoculation. Ratio of *A. gossypii* germ bubbles to spores is one third to two thirds; n=567.

Results

We characterized proteins which are required and expressed during germination by four different biological replicas of the five hour time point. The total protein extract of three biological replicas was directly analyzed in MS/MS run without prefractionation (for more details see Appendix 2). The fourth sample, analyzed in this Chapter, was separated into twelve fractions by isoelectric focusing as described in Chapter1 using an OffGel fractionator. All twelve fractions were analyzed by tandem MS/MS runs. Most importantly, isotope-labeled peptides were spiked into the this sample of tryptic peptides prior to OffGell fractionation. From each of the twelve fraction 2 µl peptide mixture was injected into the tandem mass spectrometer and separated with a retention time of 140 minutes. The relative abundances were determined by the All- and Top3-method, and the protein copy numbers were later adjusted using a regression line (iBAQ and the T3PQ method).

Characterization of germ bubbles in the spore/germ bubble mixture

We selected the 5h time point as representative germ bubble stage during the developmental cycle of *A. gossypii*. At this stage, the spore/germ bubble mixture showed different sizes of germ bubbles and morphologically unaltered spores. Since germ bubbles represent the isotropic growth phase of *A. gossypii*, it is important to quantify the cell stage composition of the 5h culture and define an average germ bubble for which the protein composition will be determined. Germ bubbles plus the two attached thin cones are called germlings. As already mentioned, the 5h culture consisted of 35% germ bubbles and 65% spores (Figure3-2). This indicates that, in contrast to isolated spores, the cells have proteomes with distinct differences depending on their developmental stage. The heterogeneity of germ bubbles is best seen by a scatter plot of the 225 measured diameters in Figure3-3A and their Gaussian distribution in Figure3-3B. Representative images of different germ bubbles are shown in Figure3-3C. The wide range from two to eight micrometers in diameter is caused by the asynchronous germination of spores. We calculated a mean diameter of 4.4 µm. In conclusion, processes like induction of efficient transcription and translation, secretory vesicle formation and transport, efficient cell wall synthesis, DNA duplication and many others are most likely timed in an individual spore but occur asynchronously in different spores.

Next, an aliquot of the 5h culture was stained with DAPI to visualize and compare numbers of nuclei per germ bubble. Figure3-4A shows two examples of overlayed DIC- and DAPI-stained fluorescence images with two nuclei and one nucleus, respectively. The blue stained zones represent nuclear DNA. Up to seven nuclei can be seen in Figure3-4B. Uneven numbers of nuclei confirm asynchronous mitoses in germ bubbles. In total, nuclei were counted in 147 germ bubbles and their distribution is presented as box plot in Figure3-5A. The mean was calculated as 2.5 nuclei per germ bubble. The frequency distribution in Figure3-5B shows that over 80% of the germ bubbles carry one to three nuclei with 40% carrying two nuclei. Four nuclei were found in 11% of the germ bubbles. Five or more nuclei were only found in a minority of germ bubbles. To sum up, the average germ bubble in the 5h culture has a diameter of 4.4 µm and carries 2.5 nuclei.

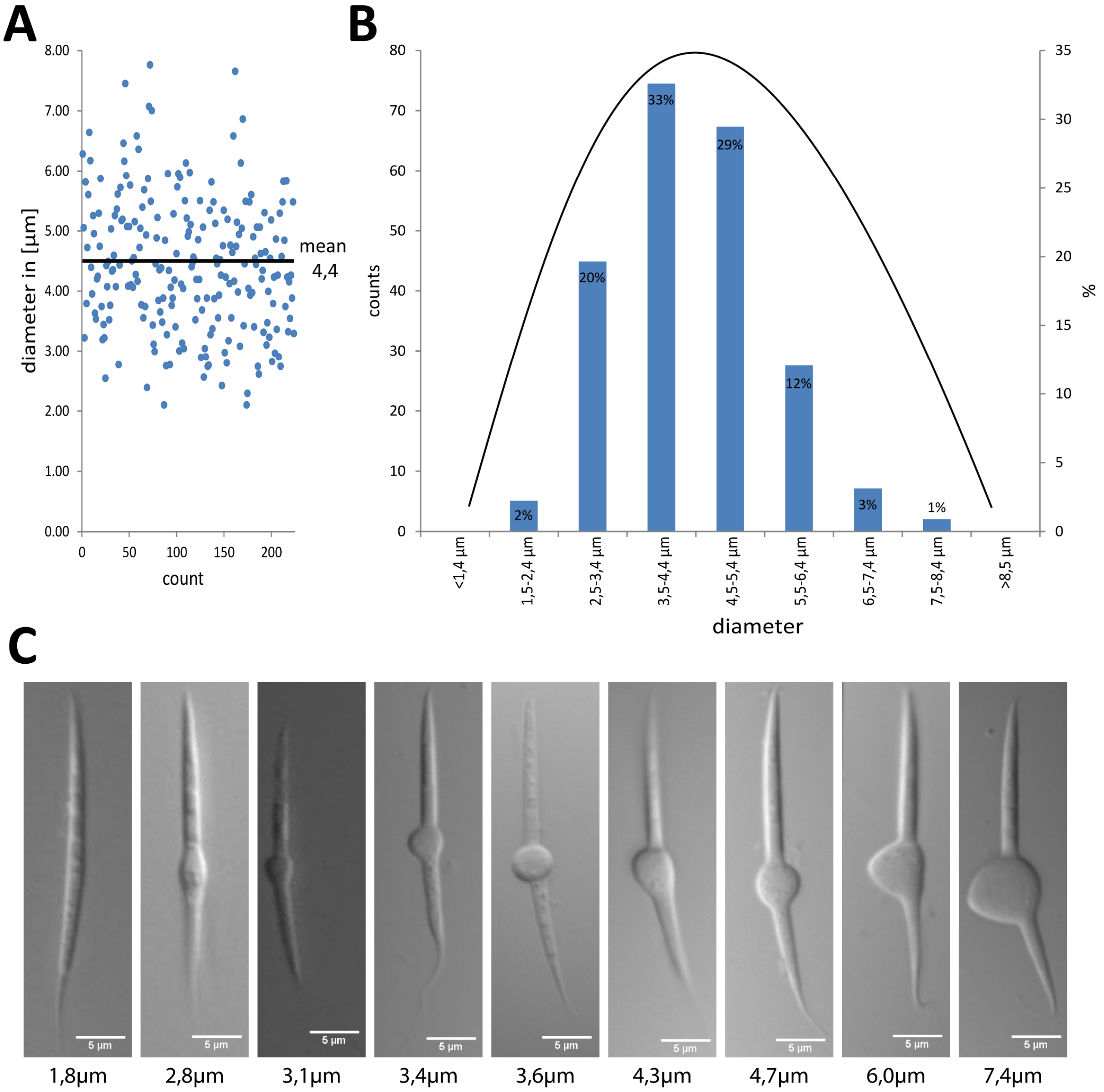


Figure3-3: Heterogeneity of germ bubble sizes 5 hours after spore inoculation. **(A)** Scatter plot of diameters of germ bubbles, n=225. Minimum size of a barely visible bubble was 2 µm and maximal size was close to 8 µm. **(B)** Gaussian distribution of germ bubble diameters. Mean is 4,4 µm. **(C)** Examples of microscopic images of different germ bubble sizes 5 hours after spore germination. Scale bar 5 µm.

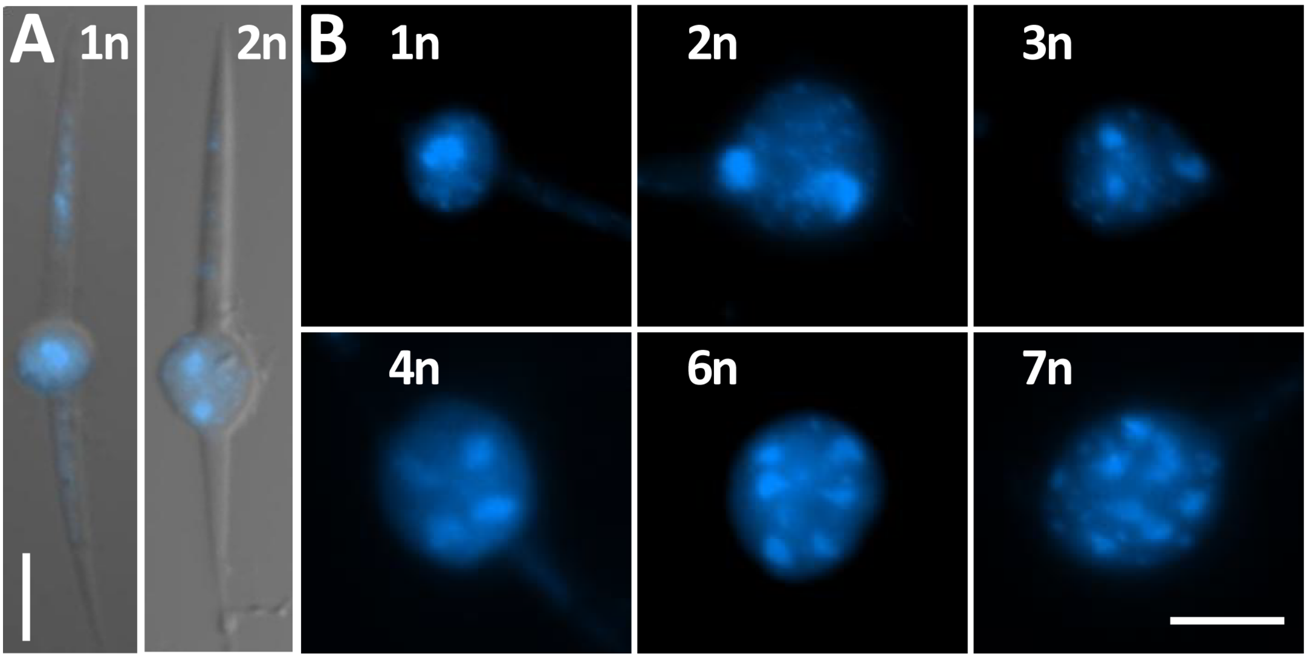


Figure3-4: Numbers of nuclei per germ bubble 5 hours after spore inoculation. An aliquot of the germ bubble/spore mixture was fixed with with 2% PFA and stained with DAPI. **(A)** Overlay of a DIC and fluorescence images of two selected germlings. Scale bar 5 µm. **(B)** Fluorescence images of germ bubbles with different numbers of nuclei. Scale bar 5 µm.

Protein abundances in an *A. gossypii* spore/germ bubble mixture 5h after spore incubation

Proteins were extracted from a known number of cells in the 5h culture and processed as described in Chapter2. Nearly 3,800 different proteins were identified in the spore/germ bubble mixture and their relative abundances were estimated using he All- and Top3 method as described for the spore proteome in Chapter2. The protein composition reflects the sum of the proteome in isotropically growing germlings, pre-germinating spores and spores. However, proteins found in higher abundances than in spores are most likely expressed during germination and very likely represent the set of proteins needed for the initiation of germination and isotropic growth of the germ bubbles. The ranking of the 20 most abundant proteins is listed in Table3-1 and their potential functions are mentioned in Table3-2. The relative abundances of all detected proteins are listed in Table3-4 (see Appendix 6). The 20 most abundant proteins can be assigned to eight functional groups:

1. Cell wall and membrane (Om45 and Cwp1)
2. Translation (Tef1)
3. Protein assembly and aggregation (Hps26, Cpr1, Tfs1)
4. Carbon metabolism (Tdh3, Fba1, Ach1 and Ltp1)
5. Redox homeostasis and protein protection against oxidative stress (Sod1, Ypr1, Tsa1)
6. Mitochondrial (Fmp45, Pet9 and Atp2)
7. Cytoskeleton (Act1)
8. Ribosomal proteins (Rps29 and Rpl2)

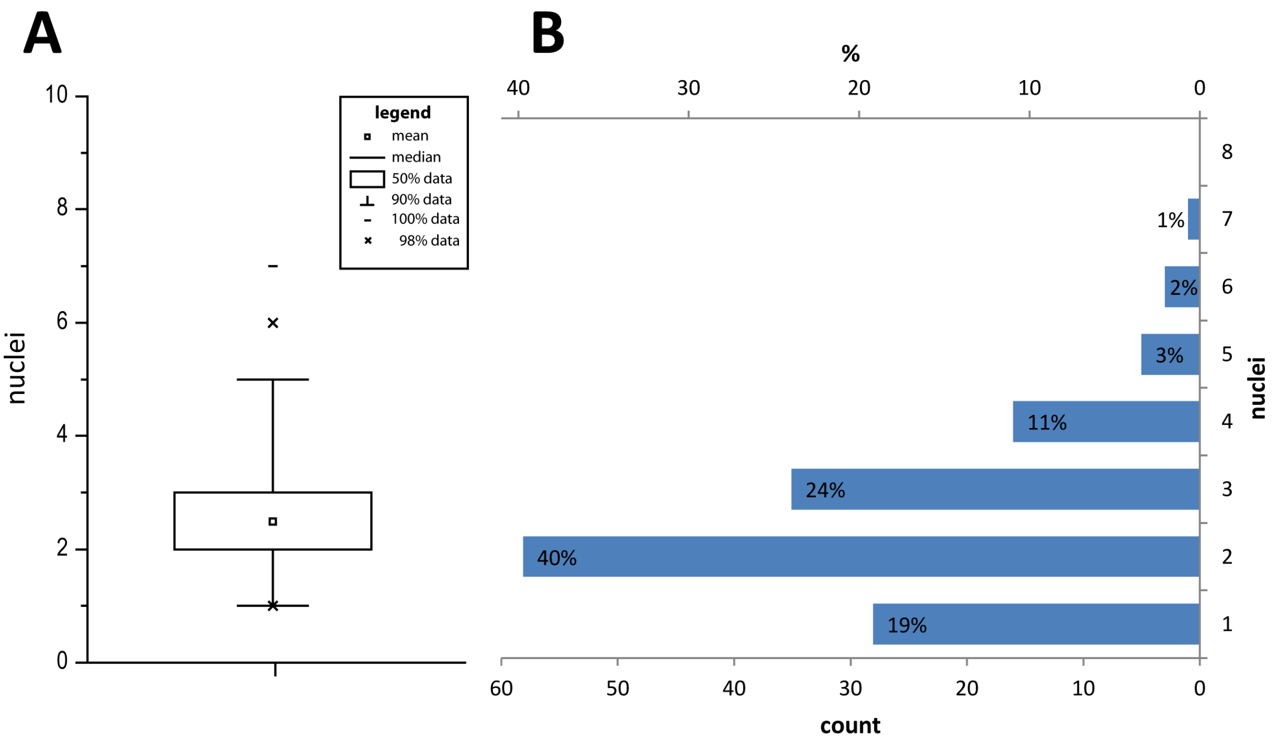


Figure3-5: Distribution of germ bubbles with different numbers of nuclei 5 hours after spore inoculation. An aliquot of the germ bubble/spore mixture was fixed with 2% PFA and stained with DAPI to count the fluorescence signals in germ bubbles corresponding to nuclei. **(A)** Box chart diagram of the distribution of germ bubbles carrying a specific number of nuclei. The mean is 2.5 nuclei per germ bubble. **(B)** Percent of germ bubbles with a specific number of nuclei. n= 147 germ bubbles.

Table3-1: Ranking of the 20 most abundant proteins in the cell mixture five hours after spore inoculation. (*1) All detected peptides are taken into account and were divided by all possible peptides (*2) Average of the top3 abundant peptides.

no.	accession	common name	5h relative abundance (All *1)	5h relative abundance (Top3 *2)	detected peptieds	possible peptides	protein size [aa]	protein size [MW]	Score
1	ADL370C	Tef1, Tef2	59134	195965	66	27	458	49876	3384
2	AAL128C	Hsp26	53014	148882	15	9	178	20199	838
3	AER052W	Rps29A, Rps29B	30665	40586	5	4	56	6605	231
4	AFR694W	YLR179C, Tfs1	26686	67206	15	8	204	22333	947
5	ABL156C	Fmp45	22814	118330	26	24	429	47034	1697
6	ADL154C	NOHBY412	21189	80132	20	14	197	21035	1109
7	ABR028C	Cwp1	19216	65634	13	12	763	70107	1138
8	AER312W	Tsa1, Tsa2	18440	50321	18	13	197	21486	1076
9	ABR068C	Fba1	16659	68804	35	17	360	39319	2287
10	AER031C	Tdh3	15574	53899	43	20	331	35479	2601
11	ADL127C	Rpl2A, Rpl2B	14927	60540	18	15	254	27320	939
12	ABR222W	Act1	14834	56527	40	24	376	41688	2355
13	AER428W	Om45	10627	57693	60	32	498	53827	3587
14	AGL177C	Cpr1	10534	25796	17	11	162	17384	1077
15	ACL140C	Rps0A, Rps0B	10399	50431	14	16	253	27957	756
16	AGL321W	Sod1	10131	19560	13	7	154	15856	747
17	ABR169W	Atp2	9711	34337	50	33	505	53992	3092
18	AFR512W	Lpd1	9647	73041	58	33	496	53332	3526
19	AER184W	Pet9, Aac3	9485	32582	33	20	305	33110	1886
20	AFR020W	Ach1	9287	39982	51	32	523	57754	2760

Interestingly, by comparing the top 20 proteins of the 5h time point with the spore proteome (Chapter2), some proteins have dramatically changed their rank. The four proteins Aco1, Ilv5, Cit1, and Hxk1/2 present in the top 20 of the spore proteome are markedly decreased by abundance after 5h and rank among the top 50 proteins. Importantly, both ribosomal proteins Rsp29 and Rpl2 are increased in abundance and reach the top 20. In particular, Rsp29 increased from rank 25 to 3 and Rpl2 from rank 36 to 8. In general, all ribosomal protein significantly increased their protein abundance in the spore/germ bubble mixture.

Protein copy numbers per nucleus of the *A. gossypii* spore/germ bubble mixture

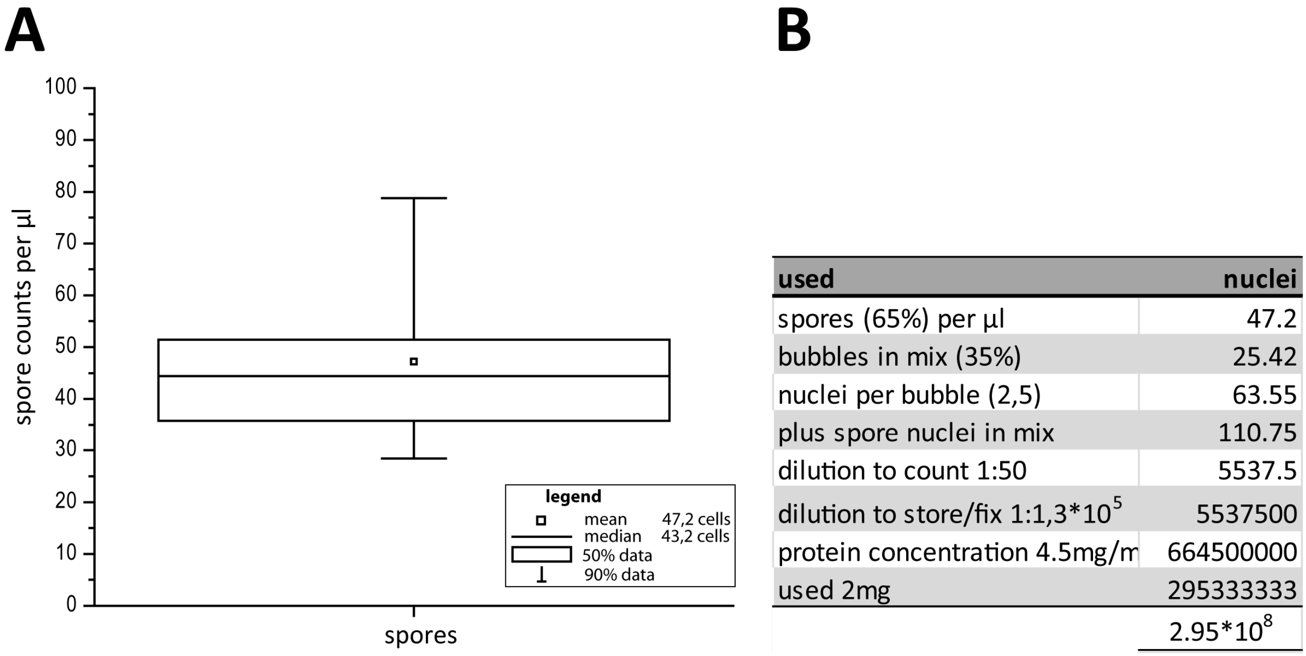
Absolute quantification of selected proteins using AQUA peptide standards

We determined protein copy numbers per nucleus through stable isotope-labeled peptides quantified by multiple reaction monitoring (MRM). Since we used tryptic peptides, one arginine or lysine residue is present at the C-terminus of each synthesized peptide. Both amino acids are labeled with carbon 13 and nitrogen 15 isotopes which shifts the mass of the peptides by 10 dalton or 8 dalton, respectively, compared to the endogenous peptides. We used 40 AQUA peptides, which were spiked into the tryptic peptide mixture with amounts of 2 or 20 pmol. The tryptic peptide mixture had a protein concentration of 2 mg/ml extracted from a known number of cells. This allowed us to determine the protein copy numbers per nucleus.

Table3-2: *Saccharomyces* Genome Database description of the syntenic homolog of *S. cerevisiae* for the top 20 highest ranked proteins by their abundance in *A. gossypii* at the five hour time point.

no.	accession	common name	name description	SGD description of the firth *S. cerevisiae* homolog
1	ADL370C	Tef1, Tef2	Translation Elongation Factor	Translational elongation factor EF-1 alpha; also encoded by TEF2; functions in the binding reaction of aminoacyl-tRNA (AA-tRNA) to ribosomes; may also have a role in tRNA re-export from the nucleus
2	AAL128C	Hsp26	Heat Shock Protein	suppress unfolded proteins aggregation; oligomer activation requires heat-induced conformational change; also has mRNA binding activity
3	AER052W	Rps29A, Rps29B	Ribosomal Protein of the Small subunit	Protein component of the small (40S) ribosomal subunit; nearly identical to Rps29Bp and has similarity to rat S29 and E. coli S14 ribosomal proteins
4	AFR694W	YLR179C, Tfs1	cdc25 (Twenty-Five) Suppressor	protein (PEBP) family member; targets to vacuolar membranes during stationary phase; acetylated by NatB N-terminal acetyltransferase
5	ABL156C	Fmp45	Found in Mitochondrial Proteome	Integral membrane protein localized to mitochondria (untagged protein); required for sporulation and maintaining sphingolipid content; has sequence similarity to SUR7 and YNL194C
6	ADL154C	NOHBY412	No homolog in Saccharomyces cerevisiae	
7	ABR028C	Cwp1	Cell Wall Protein	Cell wall mannoprotein that localizes specifically to birth scars of daughter cells, linked to a beta-1,3- and beta-1,6-glucan heteropolymer through a phosphodiester bond; required for propionic acid resistance
8	AER312W	Tsa1, Tsa2	Thiol-Specific Antioxidant	associates to form a high-molecular weight chaperone complex under oxidative stress; deletion results in mutator phenotype
9	ABR068C	Fba1	Fructose 1,6-bisphosphate aldolase	fructose 1,6 bisphosphate to glyceraldehyde-3-P and dihydroxyacetone-P; locates to mitochondrial outer surface upon oxidative stress
10	AER031C	Tdh3	Triose-phosphate DeHydrogenase	tetramer that catalyzes the reaction of glyceraldehyde-3-phosphate to 1,3 bis-phosphoglycerate; detected in the cytoplasm and cell wall
11	ADL127C	Rpl2A, Rpl2B	Ribosomal Protein of the Large subunit	Protein component of the large (60S) ribosomal subunit, identical to Rpl2Bp and has similarity to E. coli L2 and rat L8 ribosomal proteins
12	ABR222W	Act1	ACTin	Actin, structural protein involved in cell polarization, endocytosis, and other cytoskeletal functions
13	AER428W	Om45	Outer Membrane	Protein of unknown function, major constituent of the mitochondrial outer membrane; located on the outer (cytosolic) face of the outer membrane
14	AGL177C	Cpr1	Cyclosporin A-sensitive Proline Rotamase	Cytoplasmic peptidyl-prolyl cis-trans isomerase (cyclophilin), catalyzes the cis-trans isomerization of peptide bonds N-terminal to proline residues; binds the drug cyclosporin A
15	ACL140C	Rps0A, Rps0B		
16	AGL321W	Sod1	SuperOxide Dismutase	Cytosolic copper-zinc superoxide dismutase; some mutations are analogous to those that cause ALS (amyotrophic lateral sclerosis) in humans
17	ABR169W	Atp2	ATP synthase	Beta subunit of the F1 sector of mitochondrial F1F0 ATP synthase, which is a large, evolutionarily conserved enzyme complex required for ATP synthesis; phosphorylated
18	AFR512W	Lpd1	LiPoamide Dehydrogenase	Dihydrolipoamide dehydrogenase, the lipoamide dehydrogenase component (E3) of the pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase multi-enzyme complexes
19	AER184W	Pet9, Aac3	PETite colonies	synthesized ATP; also imports heme and ATP; phosphorylated; required for viability in many lab strains that carry a sal1 mutation
20	AFR020W	Ach1	Acetyl CoA Hydrolase	acetyl-CoA-hydrolase activity; phosphorylated; required for acetate utilization and for diploid pseudohyphal growth

To determine the number of cells used for the extraction of proteins, we counted spores and germ bubbles in the 5h sample. We used a similar method as described in Chapter2-C (Figure2-17). We fixed a small amount of the spore/germ bubble sample, diluted the sample and spread 5 µl each on an agar plates with a spreader rod to ensure good separation. The spreading of the cells created mechanical forces, which destroyed some of the germ bubbles. Thus, we were only able to accurately count all spores (Figure3-6A) which represent 65% of the spore/spore bubble mixture. In total, we counted an average of 47 spores per µl. To calculate the total number of cells and finally the number of nuclei, we relied on the ratio of spores to germ bubbles (65 to 35) already determined above (Figure 3-2). Applying this to the spore number, 47 spores per µl represent 65% of total cell. Therefore, 25 germ bubbles should be present per µl. We determined a mean of 2.5 nuclei per germ bubble (Figure3-5). We multiplied the 25 germ bubbles per µl with 2.5 which results together with the spore nuclei in 111 nuclei per µl of the diluted spore/germ bubble mixture. Multiplying with the different dilution steps, we concluded that 3x10^8 nuclei correspond to 2 mg protein extracted from the spore/germ bubble mixture. The same mix of AQUA peptides used for the absolute quantification of spore proteins in Chapter2-C was spiked into the tryptic peptides obtained from 2 mg protein extract of the spore/germ bubble mixture.



used	nuclei
spores (65%) per µl	47.2
bubbles in mix (35%)	25.42
nuclei per bubble (2,5)	63.55
plus spore nuclei in mix	110.75
dilution to count 1:50	5537.5
dilution to store/fix 1:1,3*10^5	5537500
protein concentration 4.5mg/m	664500000
used 2mg	295333333
	2.95*10^8

Figure3-6: Steps to determine the number of nuclei in the germ bubbles and spores used to extract 2 mg protein. **(A)** Box-plot of the number of spores per µl of 20 aliquots of a known dilution of the germ bubble/spore mixture. 5 µl aliquots were spread on 20 different plates and spores were counted in microscopic image (see Figure2-17). On average 47 spores were counted per µl. **(B)** Rationale to estimate the number of cells in the germ bubble/spore mixture, expressed as number of nuclei, used for isolating 2mg protein.

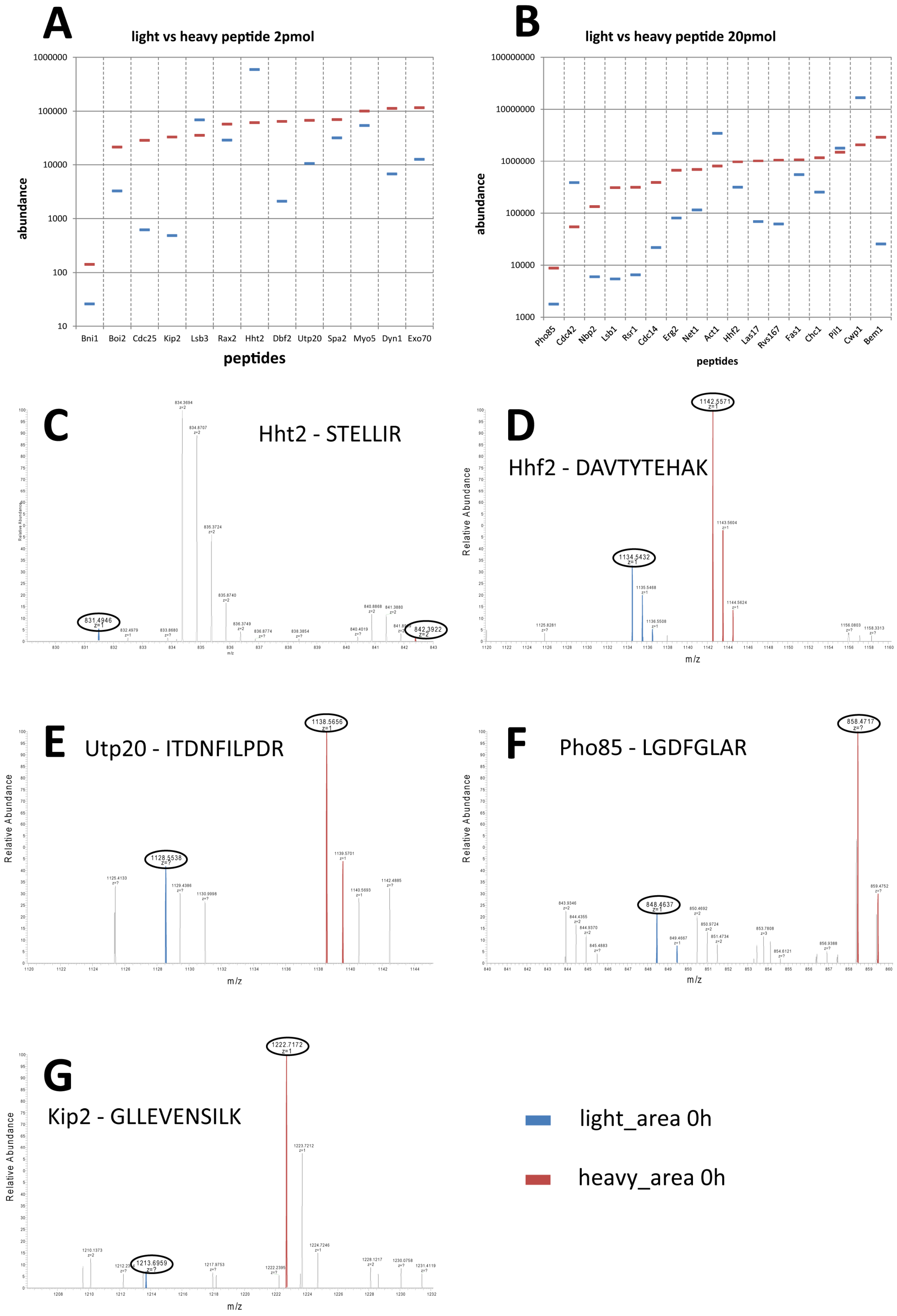

A
light vs heavy peptide 2pmol
abundance
1000000
100000
10000
1000
100
10
Bni1
Boi2
Cdc25
Kip2
Lsb3
Rax2
Hht2
Dbf2
Utp20
Spa2
Myo5
Dyn1
Exo70
peptides
B
light vs heavy peptide 20pmol
abundance
100000000
10000000
1000000
100000
10000
1000
Pho85
Cdc42
Nbp2
Lsb1
Rsr1
Cdc14
Erg2
Net1
Act1
Hhf2
Las17
Rvs167
Fas1
Chc1
Pil1
Cwp1
Bem1
peptides
C
Hht2 - STELLIR
831.4946 z=1
842.3922 z=2
834.3694 z=2
834.8707 z=2
835.3724 z=2
835.8740 z=2
m/z
D
Hhf2 - DAVTYTEHAK
1142.5571 z=1
1134.5432 z=1
1143.5604 z=1
1135.5468 z=1
1144.5624 z=1
Relative Abundance
m/z
E
Utp20 - ITDNFILPDR
1138.5656 z=1
1128.5538 z=?
1139.5701 z=1
Relative Abundance
m/z
F
Pho85 - LGDFGLAR
858.4717 z=?
848.4637 z=1
859.4752 z=?
Relative Abundance
m/z
G
Kip2 - GLLEVENSILK
1222.7172 z=1
1213.6959 z=?
1223.7212 z=1
1224.7246 z=1
Relative Abundance
m/z
light_area 0h
heavy_area 0h

Figure3-7: Abundances of 30 pairs of heavy/light peptides in MS1-spectra for analyzing germ bubble/spore proteins. **(A)** Red bars: Abundances of 13 AQUA peptides added as 2pmol. Blue bars: Abundances of the corresponding light peptides. **(B)** Red bars: Abundances of 17 AQUA peptides added as 20 pmol. Blue bars: Abundances of the corresponding light peptides. The proteins from which the reference peptides were selected are indicated at the bottom of the graphs. **(C to G)** Examples for sections in MS1-spectra showing heavy and light peptides. All peptide ions are singly-charged, with a mass shift of 1 dalton. Peptide peaks matching endogenous peptides are colored in blue and the heavy isotope-labeled homologues in red. **(C)** The Hht2 peptide has a calculated mass of 830.4861 dalton and the heavy peptide is shifted by 10 dalton. The heavy peptide was spiked into the sample with 2 pmol and was detected in fraction 4 at 29.03 min with a maximal MS1-Spectrum intensity of $4.16*10^7$. The ratio between light and heavy peptide is 9.7. **(D)** The Hhf2 peptide has a calculated mass of 1133.5353 dalton and, which is shifted by 8 dalton for the heavy peptide. The heavy peptide has a concentration of 20 pmol and was detected in fraction 4 at 11.36 minutes with a maximal MS1-Spectrum intensity of $3.08*10^6$. The ratio between heavy and light is 3.1. **(E)** The Utp20 peptide has a calculated mass of 1127.5472 dalton and the heavy peptide has an increased mass by 10 dalton. With 2 pmol the heavy peptide was spiked into the tryptic peptide mix. The heavy peptide was detected in fraction 7 at 22.48 min with a maximal MS1-Spectrum intensity of $1.21*10^5$. The ratio between heavy and light is 6.3. **(F)** The Pho85 peptide has a calculated mass of 847.4552 dalton and the heavy peptide has a mass of 857.45Da. The heave peptide was spiked into the sample with a concentration of 20 pmol and was detected in fraction 6 at 38.00 minutes with a maximal MS1-Spectrum intensity of $5.22*10^6$. The ratio between heavy and light peptide is about 5. **(G)** The endogenous peptide of Kip2 has a mass of 1213.705 dalton and 8 dalton more for the heavy peptide. The heavy peptide has a concentration of 2 pmol and was detected in fraction 2 at 68.39 min with a maximal MS1-Spectrum intensity of $1.21*10^6$. The ratio between heavy and light is about 69.

First, we absolutely quantified 30 reference proteins for which AQUA peptides were detected by comparing the intensities of the MS1 peaks of the endogenous peptides and the isotope-labeled peptides (Figure3-7). The abundances of light peptides were marked with blue bars and abundances of the heavy peptides with red bars. Because of the large intensity differences by several orders of magnitude for the detected peptides the abundances are plotted in a logarithmic scale (Figure3-7A and B). For example, the range of heavy peptide abundances was 142 to 2,905,400. High resolution MS1 spectra are shown for five pairs of light and heavy peptides to determine absolute abundances (Figure3-7 C to G). Peptide ions are charged by protonation and shift in mass, so that the theoretical ion mass is not identical with the measured mass. Figure3-7C shows a MS1-spectrum of the fourth fraction after 29.03 minutes between 830 and 843 dalton. The heavy and light peptide STELLIR representing the histon Hht2 are marked. Both peaks are barely visible and show a low intensity compared to the highest peak in this MS1-spectrum. The relative abundance of the light peptide is about 9 times higher than the heavy peptide. The peptide DAVTYTEHAK of the Hhf2 protein is found in the fourth fraction and at 11.36 minutes (Figure3-7D). The peptide sequence carries 10 amino acids with one lysine. By comparing the light and the heavy peptide abundance, the mass shifts by 8 dalton from 1134.54 to 1142.56 dalton, because of the C13 and N15 isotope-labeled lysine. The heavy peptide differs by a factor 3.1 compared to the light peptide. Figure3-7E to G show a smaller abundance of all endogenous peptides compared with the corresponding heavy peptides. The mass of the Utp20 peptide shifted from 1128.56 to 1138.56 dalton and the mass of the Pho85 shifted by 10 dalton.

Table3-3: Protein copies per nucleus calculation with the measured abundance based on the ratio between the AQUA peptide and the endogenous peptide. Different steps and results of the calculation are present. Additionally column 13-15 transmit the logarithmic definition of the abundance and the cpn to design the regression lines (figure3-8).

accession	common name	relativ abundance		ratio from area	heavy peptide amount in	concentration in [pm]	nucleus count; 5h post inoculation	protein concentration in pmol per	**copies per nucleus**	protein abundance / All-method	protein abundance / Top3-method	log copies per nucleus	log protein abundance / All-method	log protein abundance / Top3-method
		light area	heavy area											
ABR028C	Cwp1	16720400	2081205	8.0340	20	160.6799	2.95E+08	5.44E-07	**327635**	19216	65634	5.515	8.284	8.817
AGL093W	Cdc42	386355	54416	7.1000	20	142.0005	2.95E+08	4.81E-07	**289546**	1002	1816	*5.462*	*7.001*	*7.259*
ABR222W	Act1	3462054	801692	4.3184	20	86.3687	2.95E+08	2.92E-07	**176110**	14834	56527	5.246	8.171	8.752
AEL329W	Pil1	1799185	1504350	1.1960	20	23.9198	2.95E+08	8.10E-08	**48774**	3550	14303	4.688	7.550	8.155
ADL202C	Hht2	593320	61175	9.6987	2	19.3975	2.95E+08	6.57E-08	**39552**	252	379	*4.597*	6.401	*6.578*
AER085C	Fas1	547530	1066610	0.5133	20	10.2667	2.95E+08	3.48E-08	**20934**	1672	22388	4.321	7.223	8.350
ADL201W	Hhf2	315744	972942	0.3245	20	6.4905	2.95E+08	2.20E-08	**13234**	2630	4570	4.122	7.420	7.660
AER359W	Chc1	253400	1176000	0.2155	20	4.3095	2.95E+08	1.46E-08	**8787**	409	3539	3.944	6.612	7.549
AGL242C	Pho85	1780	8760	0.2032	20	4.0639	2.95E+08	1.38E-08	**8287**	90	310	3.918	*5.955*	*6.491*
AEL017W	Lsb3	69005	35688	1.9336	2	3.8671	2.95E+08	1.31E-08	**7885**	103	576	3.897	6.012	6.761
AAL181C	Net1	114903	690760	0.1663	20	3.3269	2.95E+08	1.13E-08	**6784**	53	777	3.831	5.728	6.890
AFL181C	Erg2	80438	665553	0.1209	20	2.4172	2.95E+08	8.18E-09	**4929**	270	535	3.693	6.431	6.728
AGR285W	Las17	68845	1014400	0.0679	20	1.3574	2.95E+08	4.60E-09	**2768**	30	209	3.442	5.479	6.321
AFR140C	Rvs167	61857	1050050	0.0589	20	1.1782	2.95E+08	3.99E-09	**2402**	179	665	3.381	6.252	6.823
AEL025W	Cdc14	21833	388146	0.0563	20	1.1250	2.95E+08	3.81E-09	**2294**	92	631	3.361	5.964	6.800
AEL306C	Myo5	54162	100776	0.5375	2	1.0749	2.95E+08	3.64E-09	**2192**	121	1847	3.341	6.081	7.267
AGR095W	Rax2	29100	57400	0.5070	2	1.0139	2.95E+08	3.43E-09	**2067**	15	156	3.315	5.167	6.192
ADL022C	Spa2	32040	69960	0.4580	2	0.9160	2.95E+08	3.10E-09	**1868**	3	109	*3.271*	*4.538*	6.036
AGL169C	Nbp2	5978	133600	0.0447	20	0.8949	2.95E+08	3.03E-09	**1825**	21	63	3.261	5.322	*5.800*
AFR464W	Rsr1	6533	313492	0.0208	20	0.4168	2.95E+08	1.41E-09	**850**	45	267	2.929	5.657	6.426
AFR669W	Bni1	26	142	0.1848	2	0.3695	2.95E+08	1.25E-09	**754**	3	81	*2.877*	*4.419*	5.907
AFR320W	Lsb1	5449	308457	0.0177	20	0.3533	2.95E+08	1.20E-09	**720**	24	69	2.858	5.376	5.840
AEL006W	Utp20	10560	67600	0.1562	2	0.3124	2.95E+08	1.06E-09	**637**	4	92	2.804	4.622	5.963
AGL293C	Boi2	3260	21600	0.1509	2	0.3019	2.95E+08	1.02E-09	**615**	21	299	2.789	5.317	6.476
AFR100W	Exo70	12800	116400	0.1100	2	0.2199	2.95E+08	7.45E-10	**448**	47	390	2.652	5.669	6.591
AEL241W	Bem1	25494	2905400	0.0088	20	0.1755	2.95E+08	5.94E-10	**358**	226	1489	*2.554*	*6.354*	*7.173*
ACR258W	Dyn1	6760	112600	0.0600	2	0.1201	2.95E+08	4.07E-10	**245**	4	203	2.389	4.585	6.308
ADR033W	Dbf2	2105	64700	0.0325	2	0.0651	2.95E+08	2.20E-10	**133**	10	125	2.123	4.993	6.095
ADL038W	Cdc25	621	28710	0.0216	2	0.0432	2.95E+08	1.46E-10	**88**	4	133	1.945	4.646	6.124
ACR145W	Kip2	486	33200	0.0146	2	0.0293	2.95E+08	9.91E-11	**60**	0.1	2	*1.776*	*2.801*	*4.326*
ABR008C	Hse1	0	0	0	2	0.00	2.95E+08	0.00	**0**	124	992	*0.000*	*6.095*	*6.996*
AEL090C	Inp54	0	0	0	2	0.00	2.95E+08	0.00	**0**	23	146	*0.000*	*5.367*	*6.166*
AFR441C	Tub1	0	0	0	20	0.00	2.95E+08	0.00	**0**	1035	3481	*0.000*	*7.015*	*7.542*
AGL364C	Bnr1	0	0	0	2	0.00	2.95E+08	0.00	**0**	0.1	2	*0.000*	*2.801*	*4.381*
AGR043W	Num1	0	0	0	2	0.00	2.95E+08	0.00	**0**	2	123	*0.000*	*4.383*	*6.091*

The majority of heavy peptides spiked with 2 pmol showing similar abundances between 26,000 and 100,000. Only the Bni1 heavy peptide was detected with a low abundance of 142. Heavy peptides spiked with 20 pmol were determined with an abundance range between ten thousand and three million. The distribution of the heavy peptides was excepted to be similar due to the same spiked concentrations. Different MS1 abundances are observed, because peptides differ in their abilities for ionization or detection. In total, only five isotope-labeled peptides were not detected. These peptides represent the proteins Bnr1, Inp45, Num1, Tub1, and Hse1. In contrast to the heavy peptides, light peptides showed higher variety in abundance. We determined light peptides abundances ranging from 26 (Bni1) to 16,720,400 (Cwp1). All exact abundances for heavy and light peptides are listed in Table3-3. For example, the determined abundance of endogenous and synthesized peptide of Kip2 considerably differ by a factor of 66. The abundance of the heavy peptide was determined with 33,000 and for the endogenous peptide only with 500.

In total, we could quantify 30 peptides and could therefore also determine the protein copy numbers per nucleus. To calculate the protein copy numbers, we first determined the concentration of each peptide by multiplying the spiked concentration of AQUA peptides (2 pmol or 20 pmol) with the ratio of the heavy and light peak area. Next, we determined the protein concentration per nucleus based on the known correlation of 2mg protein extracted from cells with a total of $3*10^8$ nuclei (Figure3-6). By multiplying this result with the Avogadro constant ($6.022*10^{23}$), we could define the protein copy numbers per nucleus. The final copy numbers per nucleus of the reference proteins are listed in Table3-3 colum10. The protein copy numbers in the spore/germ bubble mixture ranged from 60 copies per nucleus to over 300,000 copies per nucleus. The microtubule-specific kinesin motor Kip2 had the lowest number of proteins per nucleus and the wall protein Cwp1 was identified with the highest copies per nucleus. Interestingly, Cwp1 is also one of the most abundant proteins in the spore/germ bubble mixture (Table3-1).

Quantification of proteins normalized as copy numbers per nucleus

We determined the protein copy number per nucleus, since most germ bubbles carry more than one nucleus (Figure3-4 and 3-5) and a protein copy number per cell would not allow a comparison to spores or multinucleated cell stages. To determine the protein copy numbers for all identified proteins normalized to copies per nucleus, we developed a regression line (Figure3-8) based on the AQUA peptides (according to Chapter2C). First, the relative protein abundances of the 3u reference proteins estimated by the All- and Top3-method (Table3-3 column 11 and 12) were transformed into a logarithmic scale (column 14 and 15). Next, we calculated the logarithms of the protein copy numbers determined by AQUA peptides (column 13). Finally, the log of the protein copy numbers (x-axis) and the log of the relative abundances (y-axis) were used to calculate a linear fit line for both sets of data (Figure3-8). Values that are marked in light gray in Table3-3 column 13-14 were not taken into account to determine the regression lines. The linear fit results in a mathematical function shown in panels A and B of Figure3-8, the parameters of which are used to determine the copy numbers per nucleus from the relative protein abundance obtained with the All- and the Top3-method, respectively. Using the regression line in combination with the All-method, resulted in protein copies per nucleus for all indentified proteins in the sample (iBAQ). Using the regression line with the abundances according to theTop3-method, the obtained data are listed as T3PQ. The R-Square for the iBAQ method is 0.86 and for the T3PQ method 0.78, which indicates a better correlation of the iBAQ method.

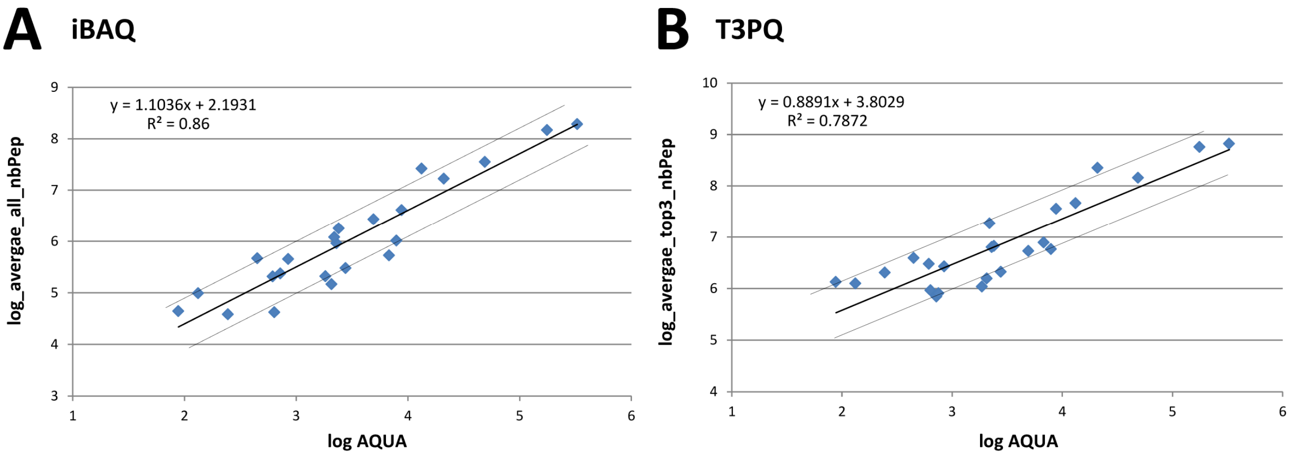


Figure3-8: Regression line to compute protein copies per nucleus via protein abundance in the 5 hour cell culture. **(A)** Plotting of the logarithm of relative abundances (y-axis) obtained with the All-method against the logarithm of the copy number per nucleus accurately determined via the heavy – light peptide ratio (x-axis) of 23 reference proteins (Table3-3). The formula for the linear fit line is shown with errors of +/- 0.09718 for 1.1036 and +/- 0.34829 for 2.19309. R^2 is 0.86 and P smaller then 0.0001. **(B)** Plotting of the logarithm of relative abundances (y-axis) obtained with the Top3-method against the logarithm of the copy number per nucleus accurately determined via the heavy – light peptide ratio (x-axis) of 24 reference proteins. . The formula for the linear fit line is also shown with errors of +/- 0.09856 for 0.88913 and +/- 0.35064 for 3.80294. R^2 is 0.7827 and P smaller then 0.0001. Thin lines in both panels show the borders of plus and minus 0.5 alterations.

The transformed formula from y (abundance) to x (cpn) based on the regression line is $\text{cpn}=10^{((\text{Log(protein abundance)}-A)/B)}$. The values for all calculated copies per nucleus are listed in Table3-4, described in Appendix 6. Only about 25 of the proteins have a copy number per nucleus which is below one. In columns 1 and 2 of this table the systematic nomenclature and the name of the proteins are listed. The relative abundance for the All-method is presented in column 3 followed by the relative abundance of the Top 3-method in column 4. The copy numbers per nucleus are listed in column 5 (iBAQ) and column 6 (T3PQ), respectively.

To compare the absolute protein copy numbers per nucleus (cpn) based on the iBAQ and T3PQ method, we summarize the results for the 35 reference proteins in Table3-5. In general, cpn numbers determined with different methods do not differ much. For example, Cwp1 is determined with 327,635 cpn based on the AQUA peptide and with the IBAQ method we calculated 349,232 cpn. Furthermore, we compared the ratio between the AQUA peptide calculations and both regression lines of T3PQ and iBAQ in Table3-5 (column 6 and 7). Together with the higher R-Square of the iBAQ method, the ratios also favor iBAC to determine copy numbers per nucleus.

Table3-5: Protein copies per nucleus in the five hour cell culture measured by AQUA peptides with computed protein copy numbers *via* the regression line. *1) All detected peptides were taken in to account and the abundance was divided by the number of all possible peptides for a protein. *2) The average of the 3 peptides with the highest abundance was used to calculate the copies per nuclei.

accession	common name	cpn (AQUA peptides)	cpn all *^{1}	cpn top3 *^{2}	ratio all *^{1}	ratio top3 *^{2}
ABR028C	Cwp1	**327635**	349232	436180	-1.07	-1.33
AGL093W	Cdc42	**289546**	24127	7716	12.00	37.53
ABR222W	Act1	**176110**	276325	368725	-1.57	-2.09
AEL329W	Pil1	**48774**	75792	78601	-1.55	-1.61
ADL202C	Hht2	**39552**	6924	1324	5.71	29.87
AER085C	Fas1	**20934**	38357	130105	-1.83	-6.22
ADL201W	Hhf2	**13234**	57783	21782	-4.37	-1.65
AER359W	Chc1	**8787**	10733	16342	-1.22	-1.86
AGL242C	Pho85	**8287**	2730	1056	3.04	7.85
AEL017W	Lsb3, Ysc84	**7885**	3074	2122	2.57	3.72
AAL181C	Net1, Tof2	**6784**	1701	2968	3.99	2.29
AFL181C	Erg2	**4929**	7366	1950	-1.49	2.53
AGR285W	Las17	**2768**	1013	680	2.73	4.07
AFR140C	Rvs167	**2402**	5073	2494	-2.11	-1.04
AEL025W	Cdc14	**2294**	2785	2351	-1.21	-1.02
AEL306C	Myo5, Myo3	**2192**	3554	7865	-1.62	-3.59
AGR095W	Rax2	**2067**	529	487	3.91	4.24
ADL022C	Spa2, Sph1	**1868**	143	324	13.06	5.77
AGL169C	Nbp2	**1825**	731	176	2.50	10.37
AFR464W	Rsr1	**850**	1467	892	-1.73	-1.05
AFR669W	Bni1	**754**	111	233	6.79	3.24
AFR320W	Lsb1, Pin3	**720**	818	196	-1.14	3.67
AEL006W	Utp20	**637**	170	269	3.75	2.37
AGL293C	Boi2, Boi1	**615**	724	1015	-1.18	-1.65
AFR100W	Exo70	**448**	1507	1366	-3.36	-3.05
AEL241W	Bem1	**358**	6274	6169	-17.53	-17.23
ACR258W	Dyn1	**245**	158	656	1.55	-2.68
ADR033W	Dbf2, Dbf20	**133**	368	379	-2.77	-2.85
ADL038W	Cdc25, Sdc25	**88**	179	408	-2.03	-4.64
ACR145W	Kip2	**60**	4	4	15.00	15.00
ABR008C	Hse1	**0**	3656	3907	0	0
AEL090C	Inp54	**0**	802	454	0	0
AFR441C	Tub1, Tub3	**0**	24849	16036	0	0
AGL364C	Bnr1	**0**	4	4	0	0
AGR043W	Num1	**0**	103	374	0	0
				mean	2.88	4.69

Adjusted abundances for a specific germ bubble proteome

The observed heterogeneous cell morphologies in cultures with germinating spores originate from asynchronous germ bubble initiations. This asynchrony prevents a specific analysis of solely germ bubble proteins, because many morphologically unchanged spores are still present 5 hours after spore inoculation (Figure3-2). Since we were able to characterize the proteome of a homogeneous spore preparation (Chapter2) we also wanted to define the germ bubble proteome by subtracting the protein content of spores from the proteins analyzed in the germ bubble/spore mixture. We used two approaches to extract the germ bubble proteome from the mixed proteome determined in the previous section. In the first approach we deducted from the mixed proteome the copy numbers of proteins that were corresponding to the 65% spores in the mixture. For the second approach, we solely looked at the volume of the germ bubbles without the attached cones the volumes of which were added to the spore fraction.

The germ bubble proteome defined by the increase in protein mass during 5 hours incubation.

We inoculated spores which consist of 5.83 mg dry biomass containing 0.31 mg protein (Figure2-15). After five hours incubation, the biomass in the spore/germ bubble mixture has increased to 9.33 mg with a protein content of 0.54 mg (Figure3-9). The biomass increased by the factor of 1.60 from 5.8 mg to 9.3 mg (Figure3-9B) and the protein content increased by a factor of 1.63 from 0.31 mg to 0.54 mg (Figure3-9C). Taken into account, that the total amount of cells were composed of one-third germ bubbles and two-thirds spores (Figure3-2), we assume that two-thirds (3.8 mg) of the original spore biomass and two thirds (0.202 mg) of the original spore protein amount belong to the spores which did not germinate during the 5 hours incubation and remained unchanged. Consequently, the one-third of spores which germinated are responsible for the increase in biomass and protein amount during the 5 hours incubation. Thus, the spore biomass of 2.0 mg grew to a germ bubble biomass of 5.5 mg (Figure3-9B) and at the same time 0.109 mg spore protein increased to 0.334 mg germ bubble protein (Figure3-9B). In conclusion, the germ bubble protein amount of 0.536 mg can be separated into 0.202 mg spore proteins and 0.334 mg germ-bubble specific proteins (Figure3-9D). Interestingly, the 35% germ bubbles account for 60% of the protein amount, whereas the 65% of spores contribute only to 40% of the protein amount after 5 hour of incubation.

Finally, we calculated the germ bubble proteome by subtracting two thirds of the spore protein copy numbers from the protein copy numbers determined for the mixed spore/germ bubble proteome compiled in Table3-6A (Appendix 6). As previously discussed, the All-method seems to generate higher quality data than the Top3 method. First, we multiplied the abundance of zero hour time point (column3) and five hours (column7) time point with the specific molecular weight (MW) (column13) of each protein. The results of the MW abundances are shown in column four and eight of Table3-6A. For each time point separately, we then summed all protein MW abundances and set these to 100 percent. For each protein, we then calculated the percentage of the total MW abundances. The 100 percent are equal to the determined protein amount in mg for each time point. In the spore/germ bubble mixture, we determined an adjusted protein amount of 0.536 mg. The total composition is unraveled in 0.334 mg proteins originating from all germ bubbles and in 0.202 mg which are extracted from all spores (Figure3-9D). The percent of each protein in the zero hour time point was multiplied by 0.202 mg spore proteins and resulted in the protein-specific spore protein amount (Table3-6A, column 5). For the five hour time point, we multiplied the percentage of each protein by the total spore/germ bubble mixture protein amount of 0.536 mg, which results in the protein specific spore/germ bubble protein amount (column 9). In the final step, we subtracted the specific spore protein amount from the specific spore/germ bubble protein amount for each protein. The result is listed in column ten and shows the amount of germ-bubble specific proteins in mg.

The germ bubble proteome defined by the ratio of the germ bubble volume and the total cell volume

For the second approach we focused on the volume composition of the spore/germ bubble mixture to extract the germ bubble-specific proteome from the mixed proteome. Since germ bubbles have different sizes and also contain non-growing compartments composed of two thin cones on opposite sides of the germ bubble (Figure3-3), we separately determined the volumes of spore bubbles and the attached cones for 64 germlings (Figure3-10A) The average volume of a germinated spore (germ bubble plus two cones) is 88.1 µm^3 (Figure3-10C). The germ bubble volumes were very heterogeneous, because of the asynchronous germination shown in Figure3-3. The germinated spore can be separated in two parts, the isotropically expanding germ bubble itself and the two non-growing cones, which represent remnants of the spore. The average volume of a germ bubble compartment is 63.8 µm^3 and the average volume of two cones is 24.3 µm^3. Both compartments represent 35% of the cells at the 5 hours time point (Figure3-2). Spores are represented with 65% and have a well-defined volume of 34.6 µm^3 (Figure2-16). For 35 germ bubbles the cell mixture contains 65 spores and 35 pairs of cones. When multiplied by the average volumes of spore bubbles (88.1 µm^3), spores (34.6 µm^3) and pairs of cones (24.3 µm^3) one can calculated that germ bubbles represent 42% and spore-specific compartments 58% of the total cell volume.

In these steps, we first defined proteins belonging to germ bubbles and second, proteins belonging to spore-specific compartments (Figure3-10E). For this approach, we used the same calculation process as previously described in the first methodic approach. In contrast to the first approach, we calculated with 42% germ bubbles and 58% spore-specific compartments, which leads to 0.355 mg proteins extracted from germ bubbles and 0.180 mg proteins from spores, considering that the 310 mg protein increased to 536 mg in five hours. These numbers were then correlated with the abundances found for the spore proteome and in the mixed spore/germ bubble proteome. Subsequently, the spore proteome was subtracted from the spore/germ bubble proteome resulting in a germ bubble specific proteome (Table3-6B, Appendix 6).

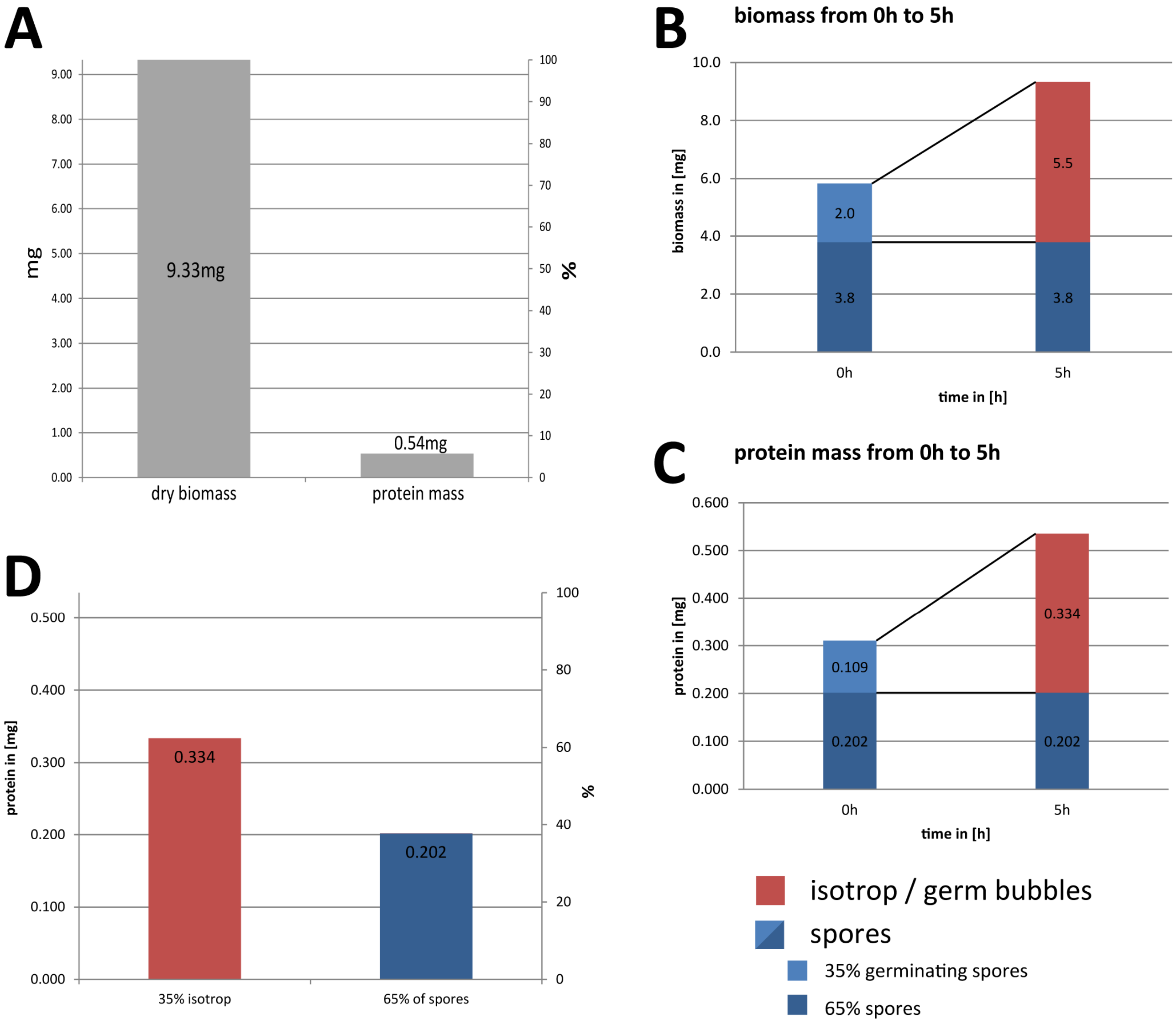
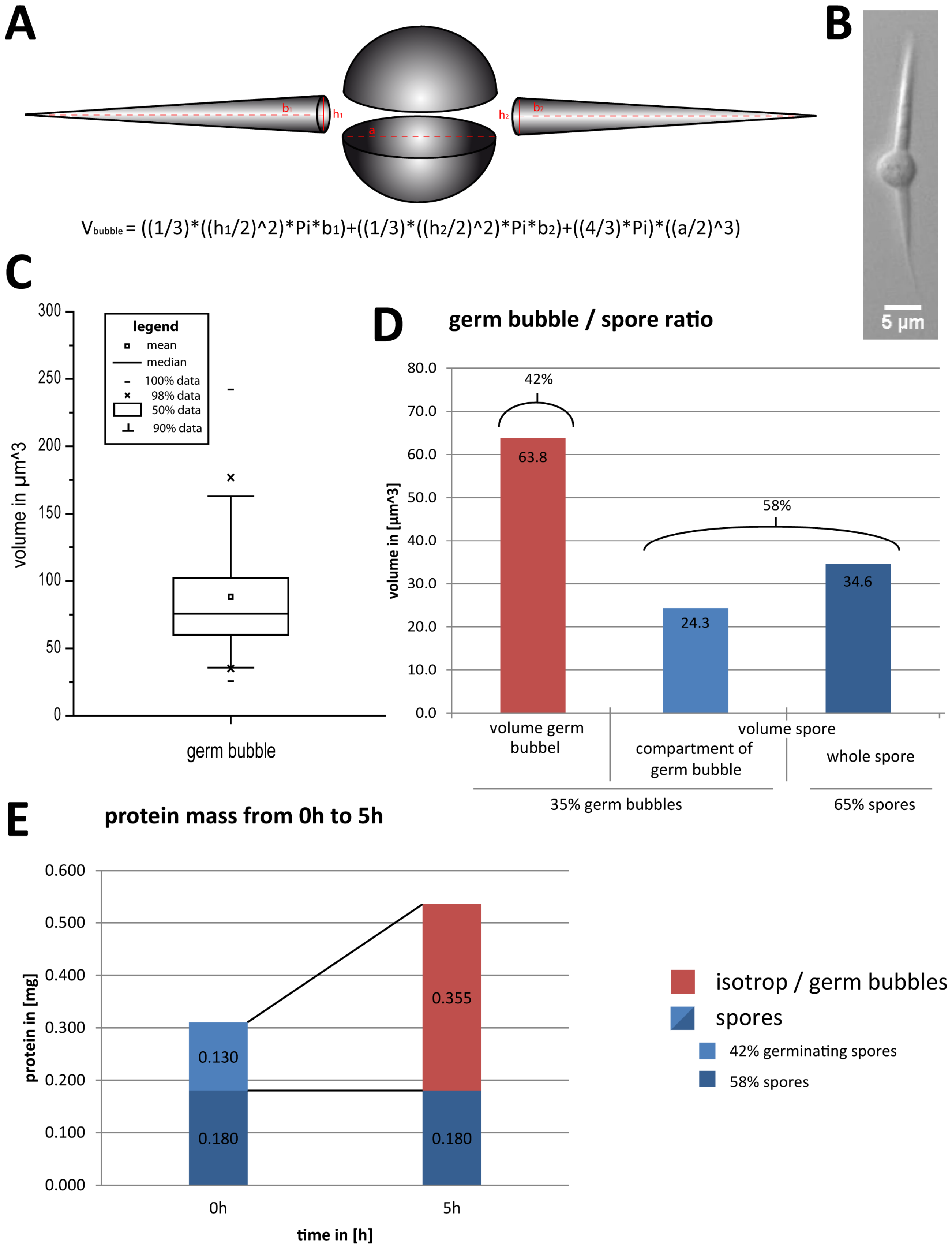


Figure3-9: Increase of biomass and protein during 5h after spore inoculation. The same number of spores analyzed in Figure2-15 was inoculated into full medium. **(A)** Correlation of dry biomass and protein amount measured in mg (y-1 axis) and specified as per cent (y-2 axis). The amount of proteins in the cell mixture is 5.8% of the biomass. **(B and C)** Increases in biomass (B) and total protein amount (C) originating from 35% germination-proficient spores indicated as light blue and red colored parts of the bar diagrams. Presumably stable contribution of the 65% non-germinating spores to the total biomass and protein amount indicated as dark blue colored parts in the bar diagrams. **(D)** Contribution of the growing germ bubbles (red) and resting spores (blue) to the total amount of protein extracted from the cell mixtures.

.



$$V_{bubble} = ((1/3)*((h_1/2)\char`^2)*Pi*b_1)+((1/3)*((h_2/2)\char`^2)*Pi*b_2)+((4/3)*Pi)*((a/2)\char`^3)$$

Protein changes from spores to germ bubbles

The distinction between germ bubble and spore proteins allows to determine protein changes during germination. Most of the 3,981 proteins identified in germ bubbles showed a positive abundance after the subtraction. With our first approach, we found 189 proteins with in a negative abundance and for the second approach only 163 proteins. These proteins resulting in negative abundance could be proteins which are no longer present in germ bubbles and are actively degraded during the germination of *A. gossypii* spores. Table3-8 lists all proteins which were found not to be present in germ bubbles. We could not quantify the 6 proteins Lsm8, Ste2, Atg19, Nsl1, Kar4, and Cep3, because peptides were only detected at the zero hour time point and not in the spore/germ bubble mixture.

Next, we characterized the protein changes from zero to five hours. The changes of protein abundances from the 0.130 mg spore proteins which are transformed into 0.355 mg germ bubble proteins are listed in Table3-7. We also used the second approach including the cell volume to achieve a more accurate profiles of abundance changes of the proteins. Next, we visualized the changes by determining the ratio between the adjusted zero hour protein abundance (column 7) and the adjusted germ bubble protein abundance (column 8). Most of the protein abundances were found to increase during the development from a spore to a germ bubble. The interpretation of the protein abundance changes based on the different development stages is described in Chapter5 “development proteomics”.

Figure3-10: Estimation of the amount of proteins in growing germ bubbles based on a volume approach. **(A)** Schematic drawing of an *A. gossypii* germ bubble. Measured distances marked as dashed red lines were used to calculate the volumes of the central germ bubble and the two attached cone compartments according to the formula shown underneath the scheme. **(B)** Microscopic DIC image of a germ bubble and the two attached thin cones which are remnants of the needle-shaped spore. Calibration bar 5µm. **(C)** Distribution of the volume of 64 germ bubbles including cones 5h after spore incubation. The mean total volume is 88.1 cubic micrometer. **(D)** Partition of the average volume of a germ bubble, the two non-growing cone compartments and a spore. **(E)** Left bar: Division of the amount of spore proteins at the time of spore inoculation into two parts, 42% will participate in generating spore bubbles during the 5h incubation (light blue) and 58% will not participate (dark blue) according to the numbers presented in C. Right bar: Division of the amount of proteins extracted from the cell mixture 5h after spore incubation into two parts, one representing the protein amount in growing germ bubbles (red) the other the remnant protein amount of the inactive spores plus the cones of the growing germ bubbles (dark blue)

Table3-8: Proteins most likely not present in germ bubbles using the normalized and adjusted relative protein amount.

accession	common name	(65%-35%) bubble proteins in 5h / in mg	(58%-42%) bubble proteins in 5h / in mg
AAL128C	Hsp26	-0.00142	-0.000126
AGL321W	Sod1	-0.000684	-0.000439
AFR312W	Nce102	-0.000172	-0.000134
ADL036C	Cda2, Cda1	-0.000132	-2.08E-05
ACR031W	Mpp10	-5.28E-05	-3.97E-05
AFL223W	Cyb5	-4.43E-05	-3.70E-05
AGR224C	NOHBY744	-3.88E-05	-3.08E-05
AGR238C	Vid28	-2.19E-05	-1.74E-05
AAL025W	Ssp120	-1.34E-05	-5.55E-06
AFR394W	Rpt3	-1.27E-05	-6.93E-06
AFR505C	YLR327C, Stf2	-1.23E-05	-9.64E-06
ACL114W	YJR107W	-1.09E-05	-7.14E-06
ADR123W	Ape3	-9.92E-06	-8.09E-07
AEL204C	Mon2	-8.91E-06	-4.26E-06
ADL077C	Mmf1, Hmf1	-8.73E-06	
ADR238C	NOHBY432	-8.06E-06	-6.99E-06
AER096C	Cnb1	-6.45E-06	-5.60E-06
ADL337W	Cmd1	-6.11E-06	-4.74E-06
AGL169C	Nbp2	-5.67E-06	-4.44E-06
ADR382W	Cyc3	-5.58E-06	-7.97E-07
AFR594W	Pbp4	-5.05E-06	-3.89E-06
AGL316W	Yuh1	-4.73E-06	-3.14E-06
AER459W	Hsp26	-3.85E-06	-2.22E-07
ACR007W	Tsr3	-2.98E-06	-2.50E-06
AGR295C	Atp14	-2.95E-06	-2.59E-06
AAL005W	Vma9	-2.79E-06	-2.23E-06
AAL089W	YOR246C	-2.72E-06	-2.40E-06
AEL340W	Sil1	-2.51E-06	-9.37E-07
AAL106W	Dap1	-2.42E-06	-2.50E-07
ABR233W	Yod1	-2.24E-06	-1.01E-06
AER354W	Erg8	-2.19E-06	
ADR231C	Stc1, Stc2	-1.84E-06	-1.40E-06
AGL155W	Taf10	-1.84E-06	-1.62E-06
AEL042C	Hxt11	-1.78E-06	-1.33E-06
AAR111C	Ahc1	-1.76E-06	-1.34E-06
AGL184W	Clf1	-1.74E-06	-1.35E-06
AMI006W	Atp6	-1.73E-06	-1.43E-06
AFR624W	Fen1, Elo1	-1.61E-06	-4.59E-07
AFL044W	Enp1	-1.41E-06	
AER162C	Dal2	-1.37E-06	-1.13E-06
AFR395C	Smp3	-1.33E-06	-1.10E-06
ADL094W	YIL024C	-1.32E-06	-1.01E-06
AFL188C	Dun1	-1.32E-06	-1.11E-06
AFR075C	Prm10	-1.30E-06	-2.59E-07
AEL262C		-1.14E-06	-6.41E-07
ADL013C	Nas2	-1.02E-06	-2.82E-07
AGL065C	Leu5	-1.00E-06	-8.97E-08
AFR706C	Ste50	-9.70E-07	-3.10E-07
ADL038W	Cdc25, Sdc25	-9.00E-07	-2.96E-09
ADR241C	Ric1	-8.71E-07	-5.01E-07
ADL027W	YAL018C	-8.50E-07	
AFR664W	Ist1	-8.22E-07	-1.86E-07
ABL151W	Dug3	-7.69E-07	-5.40E-07
AMI008W	Cob	-7.45E-07	-6.61E-07
AER270W	NOHBY526	-7.41E-07	-5.95E-07
AGR359C	Ppe1	-7.29E-07	-6.46E-07
AAL151C	Emc6	-6.88E-07	-6.12E-07
ACL125C	Lsm8	-6.65E-07	-5.92E-07
ADL246W	Vac7	-6.26E-07	-8.72E-08
AFR341C	Vps3	-6.21E-07	-3.78E-07
AER256C	Kei1	-6.04E-07	-5.21E-07
AGR165W	Ate1	-6.01E-07	-4.96E-07
AAR113W	Pns1	-5.55E-07	
ADR289C	Pho91	-5.26E-07	-3.68E-07
AGL127C	Atc1	-5.25E-07	-2.82E-07
AFR036C	YHR140W	-5.02E-07	-1.01E-07
AGR286C	YLR287C	-5.01E-07	-3.89E-07
AFR729C	Cab1	-4.89E-07	-4.16E-07
AGL011C	NOHBY701	-4.57E-07	-3.81E-07
ADR239W	Fet5	-4.56E-07	-3.19E-07
AEL154C	Dug2	-4.42E-07	-1.30E-07
AEL297W	Rad54	-4.42E-07	-3.85E-07
AGR336C	Hua1	-4.33E-07	-3.28E-07
AAR040C	Ccz1	-4.16E-07	-3.49E-07
AEL185C	Chk1	-4.04E-07	-3.44E-07
AGR225W		-4.03E-07	-3.59E-07
ADL093W	Irr1	-3.93E-07	-1.16E-07
AER431C	Vhs2, Mlf3	-3.86E-07	-3.29E-07
AFR236C	Vrg4, Hvg1	-3.49E-07	-2.78E-07
ACR008W	Dnl4	-3.47E-07	-2.80E-07
ADL218C	YNL035C	-3.40E-07	-1.40E-07
ACL088C	Rad7	-3.37E-07	-2.15E-07
AFR326W	Snq2	-3.34E-07	-2.53E-07
ABR221C	YLR065C	-3.34E-07	
ABL171W	Bre4	-3.31E-07	-2.77E-07
ADL396W	Vps75	-3.23E-07	
AAR121W	Siz1, Nfi1	-3.17E-07	-4.89E-08
AGR326C	YJR011C	-3.05E-07	-2.62E-07
AFR522C	Ste2	-3.04E-07	-2.71E-07
AEL214C	Cpd1	-2.94E-07	-9.37E-08
AAL112C	Hnt3	-2.77E-07	-1.49E-07
ADR324W	Uip5	-2.58E-07	-2.08E-07
AFR540C	YOR304C-A	-2.53E-07	-1.26E-07
ACR250W	Rpb4	-2.51E-07	-1.81E-07
AER125C	Lsm3	-2.43E-07	-2.07E-07
AFR410W	Sam4	-2.42E-07	-7.69E-08
AGR070C	Apc9	-2.42E-07	-1.74E-07
AFR365C	Ies6	-2.39E-07	-1.64E-07
ACR156W	Rny1	-2.39E-07	-2.12E-07
AFR180C	Rrn10	-2.30E-07	-2.00E-07
ABR156W	Opt1	-2.29E-07	-1.52E-07
AGR384C	Cdc43	-2.27E-07	-1.03E-07
AFL041C	Rrt2	-2.20E-07	
AFR508W	YLR050C	-2.19E-07	-1.66E-07
AGL164W	YLR143W	-2.10E-07	
AGL104C	Rav2	-2.02E-07	-9.88E-08
AER029C	Atg19	-1.91E-07	-1.70E-07
ACR227W	Skt5, Shc1	-1.86E-07	-1.31E-07
ADL368W	NOHBY420	-1.81E-07	-1.53E-07
AGL153C	YLR149C	-1.77E-07	-9.32E-08
ADR119W	YBR284W, YJL070C	-1.68E-07	-1.30E-07
ACL087C	Rho2	-1.68E-07	
AEL199W	Shr5	-1.59E-07	-1.18E-07
ACL107C	Gre3	-1.56E-07	
ABL082C	YJR141W	-1.55E-07	-6.23E-08
ADR243C	Jlp2	-1.54E-07	-1.06E-07
ADL195C	Fig4	-1.53E-07	-6.84E-08
ACL142W	Gpi1	-1.53E-07	-1.10E-07
ADR321C	NOHBY439	-1.50E-07	-1.24E-07
AER454C	YJR107W	-1.48E-07	-8.74E-08
ACL143C	Cch1	-1.48E-07	-1.04E-07
ACL138C	Sli1	-1.47E-07	-1.04E-07
AGL042W	Alg13	-1.46E-07	-6.33E-08
ABR108C	Mec1	-1.29E-07	-9.64E-08
AFR714W	Atg22	-1.28E-07	-1.01E-07
AFR187C	Irc19, Rrg4|Gon2	-1.20E-07	-9.83E-08
AFR090W	Phs1	-1.13E-07	-8.26E-08
ABL194C	Jjj1	-1.10E-07	-8.27E-08
ADL379C	Mdm36	-1.08E-07	-7.82E-08
AGR362C	Tim11	-9.74E-08	-4.08E-08
AAL004W	Ldb16	-9.22E-08	-8.18E-08
ADL002C	Inp51	-9.08E-08	
ABL008C	Vps38	-8.98E-08	-3.83E-08
AFR092W	Bck1	-8.32E-08	-6.94E-08
AAL184W	YML108W	-7.27E-08	-6.46E-08
AGL071C	Mot3	-7.07E-08	-5.64E-08
AER253W	Sem1	-6.96E-08	-5.46E-08
AFR206C	Tgl1	-6.77E-08	-3.62E-08
ADR044C	Cwc27	-5.93E-08	-3.17E-08
ABL097C	Mic14	-5.46E-08	-4.72E-08
AGR349W	Mcm21	-5.44E-08	-4.71E-08
AGR408W		-5.11E-08	-4.56E-08
AGR248C	Srb6	-5.02E-08	-3.45E-08
AFL136W	YPL230W, Rgm1	-5.01E-08	-4.40E-08
AFR728W	Kre28	-4.58E-08	-4.02E-08
AFR124W	Aps1	-4.41E-08	
ABL153W	Chs1	-4.06E-08	-2.05E-08
ADR131C	Bub2	-3.96E-08	-3.04E-08
AGR337W	Got1	-3.95E-08	-1.90E-08
AEL320W	YPR097W	-3.78E-08	
AFR432W	Ste6	-3.71E-08	-2.54E-08
AAR149W	Erv14, Erv15	-3.47E-08	
AAL016C	Bud3	-3.35E-08	
AFR604C	Spo21	-3.11E-08	-2.64E-08
AGR263C	Nej1	-3.05E-08	
AFR327C	Snu66	-2.80E-08	-2.25E-08
AGL002C	Snf8	-2.61E-08	-2.32E-08
AER328W	YML020W	-2.40E-08	-1.59E-08
AGL365C	NOHBY725	-2.39E-08	-2.12E-08
ADL212W	YNL024C	-2.12E-08	
ABR047W	Apm2	-1.98E-08	-6.32E-09
AER451W	Thi13	-1.97E-08	
AGR099C	Arp10	-1.45E-08	-1.07E-08
AER287W	Mdm1	-1.45E-08	-1.26E-08
AFR399W	Urh1	-1.36E-08	-1.38E-09
ADR051C	Dur1,2	-1.24E-08	-1.10E-08
ADR300C	Ksp1	-9.59E-09	-8.53E-09
AFL200W	Cep3	-9.44E-09	-8.41E-09
ADL333C	Uba3	-8.21E-09	-7.25E-09
AFR177C	NOHBY630	-7.85E-09	-6.50E-09
AER199C	Mrp10	-4.80E-09	-3.28E-09
AGL186C	Mtq2	-4.45E-09	-1.74E-09
AFR736C	Kar4	-3.34E-09	-2.98E-09
ADR344W	Nse1	-3.23E-09	-2.87E-09
AFR385W	Tca17	-2.90E-09	
AAR029W	Dam1	-2.83E-09	-2.47E-09
ADL260W	Oca1	-2.47E-09	
ADL393W	NOHBY421	-2.21E-09	-1.96E-09
AFL194W	Mss11	-2.04E-09	
AER246W	Ctm1	-1.49E-09	-6.09E-10
AFR284W	Sln1	-1.01E-09	-7.98E-10
ADR068W	Clb4, Clb3	-5.03E-10	-4.48E-10
AEL053C	Mms21	-3.94E-10	
AGL325W	Cps1	-3.60E-10	
AEL280W	Mlc2	-3.34E-10	-2.68E-10
ACL139W	Fre8	-9.39E-11	
AAR161W	Duo1	-6.84E-11	-5.50E-11
AFL140W	Nsl1	-2.22E-11	-1.98E-11

Discussion

The germination of the filamentous fungus *A. gossypii* is an asynchronous process that results in a heterogeneous cell mixture. After five hours of inoculation, cell cultures show 65% morphologically undifferentiated spores and 35% of germinated spores. These so-called germ bubbles vary in diameter from two to eight µm. Thus, we could not directly measure the germ bubble proteome, because spores and spore compartments were still present at five hours. To determine the germ bubble proteome, we used a direct way to estimate the composition of relative protein amounts. We subtracted the adjusted protein amounts of spores and spore compartments from the spore/germ bubble mixture, which allowed us to determine the specific germ bubble proteome. Less than 200 proteins were determined to be not present in germ bubbles. This suggests that these relative small amount of proteins are actively degraded during the first five hours. The calculation does not consider the fact that the protein composition can change also in morphologically unaffected spores. Before the spore starts to grow, cellular processes have to be activated including gene expression as well as protein degradation. Focusing on previous studies, it was shown that the mRNA levels already change after two hours (Gattiker et al., 2007). For example, the Super Oxide Dismutase Sod1, which catalyzes the bimolecular reaction from superoxide radicals to dioxygens and hydrogen peroxides (Bermingham-McDonogh et al., 1988; Steinman, 1980), is not expressed in the adjusted germ bubbles proteome. But Sod1 is an important factor for normal growth. In yeast, *sod1* null mutants exhibit mitochondrial defects including poor growth on non-fermentable carbon sources (Chang et al., 1991; Liu et al., 1992). Without the depletion of radicals that specially occur during aerobic growth, the toxicant would strongly affect the cells. For the Sod1, we argue that the old proteins, which are present in spores, are degraded because strong damages occur during the sporulation process. With the awaking of the spores during the first hours, the protein abundances have to increase. In the newly forming areas, the *de novo* synthesized proteins are only localized in a low amount in the germ bubble. Especially in the beginning of the growth on full media most of the energy can be gained from the glycolysis, and ATP is not produced via the respiratory chain, which would lead to the emergence of oxygen radicals. To gain a better understanding of the proteome, each protein has to be individually determined and analyzed, but for the specific adjustment of the abundances by hand the data volume is too large to accomplish this goal. However, when comparing the biological function of proteins and of their abundances also the changes in protein expression and localization occurring during growth have to be considered to fully understand the function and alterations in the whole proteome.

Chapter 4: Proteome of *A. gossypii* young mycelia

Chapter 4: Proteome of *A. gossypii* young mycelia

Introduction

One essential step in the *A. gossypii* life cycle is the switch from isotropic to polarized growth. When germ bubbles have reached a critical size, the first germ tube is initiated by assembly of polarity proteins perpendicular to the axis of the needle-like cones. For example, Spa2 is a component of the polarisome and a control landmark of filamentous growth. It becomes concentrated at a spot of the cell cortex which grows out into a tube, called hypha. Thus, Spa2 at outgrowing tips is a mark of polar growth (Knechtle 2002). After the first germ tube has emerged, the second germ tube is established at the opposite side of the first germ tube. Next, lateral branches are initiated along the growing hyphal cortex forming a so-called young mycelium. Nuclear divisions of *A. gossypii* occur asynchronous during the life cycle. The growth speed of *A. gossypii* hyphae accelerates over time, which leads to a rapidly increased biomass.

A culture grown for 11 hours after spore inoculation represents best the heterogeneity but also the growth potential of young mycelia. After eleven hours the developed young mycelia showed efficiently elongated hyphae and regularly established new polarity axes along the hyphal cortex. At this time point only a few non-germinated spores and germ bubbles are observed.

We specifically analyzed the proteome of young mycelia using quantitative mass spectrometry by data depending acquisition (DDA) and selected reaction monitoring (SRM) with the help of stable isotope-labeled (AQUA) peptides. In total, we determined the proteome of young mycelia grown under four conditions: A high resolution (12 fractions) analysis of an 11 hours culture, a high resolution (24 fractions) analysis from mixed cultures grown for 9 and 18 hours, and single MS/MS runs each for biological triplicates of 7 hours and 9 hours cultures. This Chapter will focus only on the high resolution quantitative proteome analysis of young mycelia present in an 11 hours culture.

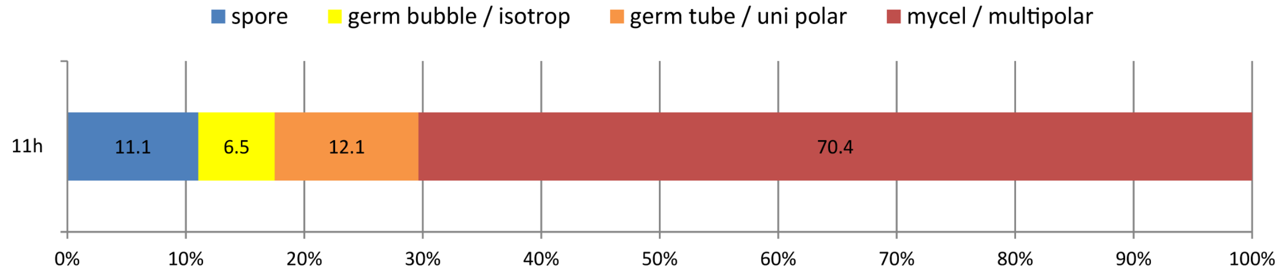


Figure4-1: Cell types seen in the culture of young mycelia 11 hours after spore inoculation. Cells were morphologically differentiated between spores, germ bubbles, germ tubes, and young mycelia. N=651.

Results

Cell composition of an *A. gossypii* culture 11 hours after spore inoculation

The majority of cells in an 11 hours *A. gossypii* culture consist of young mycelia of different morphological stages (Figure4-1). We counted 70.4% young mycelia with 2 to 20 polarly growing hyphal tips, 12.1% germlings with one germ tube, 6.5% germ bubbles, and 11.1% morphologically unchanged spores (N=651). In this overview the few germlings with two or three hyphal tubes were added to the group of young mycelia. Three different morphological stages of strictly polarly growing young mycelia are shown in Figure4-2. The development of *A. gossypii* germlings slightly differs between liquid medium and growth on agar plate. In liquid medium, germlings can develop more than two germ tubes due to the three dimensional space (Figure4-2A). A young mycelium with three established germ tubes, two lateral branches, and one emerging branch is shown in Figure4-2B. Another young mycelium in the same culture has already developed nine branches at the three main hyphae, the former germ tubes (Figure4-2C). The former germ bubble, the origin of the hyphae, can easily be identified in these image by the still attached needle-like cones.

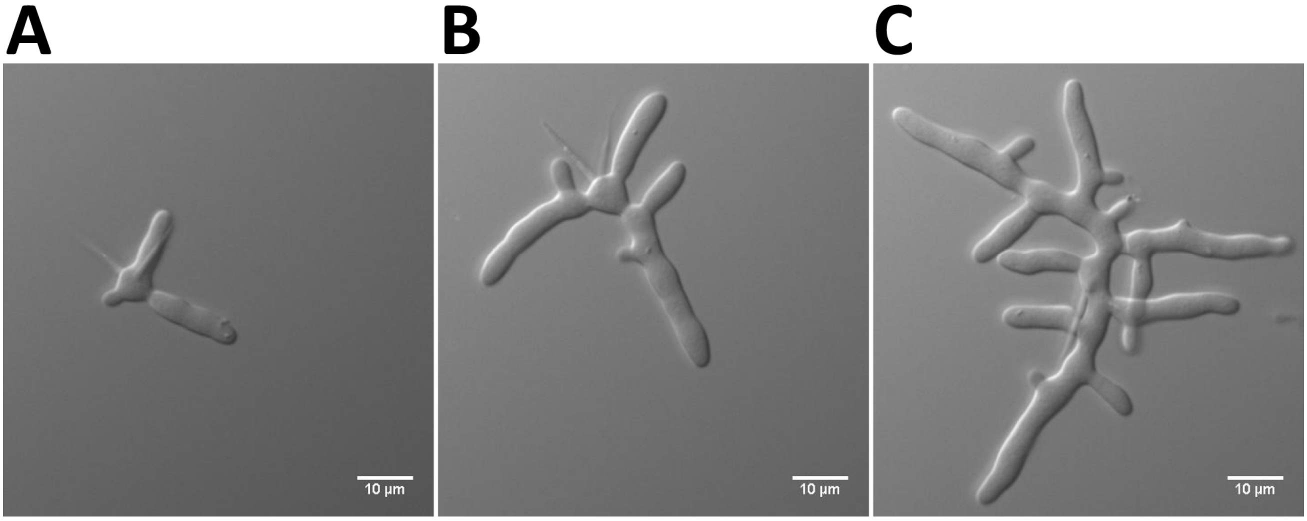


Figure4-2: DIC images of different stages of emerging hyphae and young mycelia observed after 11 hours growth in liquid full medium at 30°C. Scale bar 10 µm. **(A)** Germling with 3 germ tubes. **(B)** Very young mycelium with three hyphae, two lateral branches and one emerging branch. **(C)** Young mycelium with 14 growing hyphal tips. The needle-shaped cones mark the position of the germ bubble.

We closely looked at the diversity of the young mycelia. In particular we analyzed all polarly growing cells (including germlings with one hypha) which represent 82.5% of cells in the 11 hours culture. We determined the number of tips per young mycelium to morphologically characterize this cell stage (Figure4-3). In total, we analyzed about 200 young mycelia and detected a wide distribution from one to 23 polarized growth axes. On average, we determined about 5 tips (Figure4-3A). Next we split the box plot in a bar diagram to visualize the distribution of young mycelia based on the number of growing tips (Figure4-3B). Over 40% of the cells were still lacking lateral branches, an indication of a late start of the germination process; 13% were monopolar germlings (one germ tube), 20% were bipolar germlings (two germ tubes), and 10% had three growing tips, the majority had formed a third germ tube. Young mycelia with four to eight growing tips were each present in 7% to 10%. Also mycelia with nine and more tips were frequently seen and amounted to 12% of the 192 mycelial cells. The furthest developed mycelium carried 23 growing tips. However the number average is five tips per young mycelium.

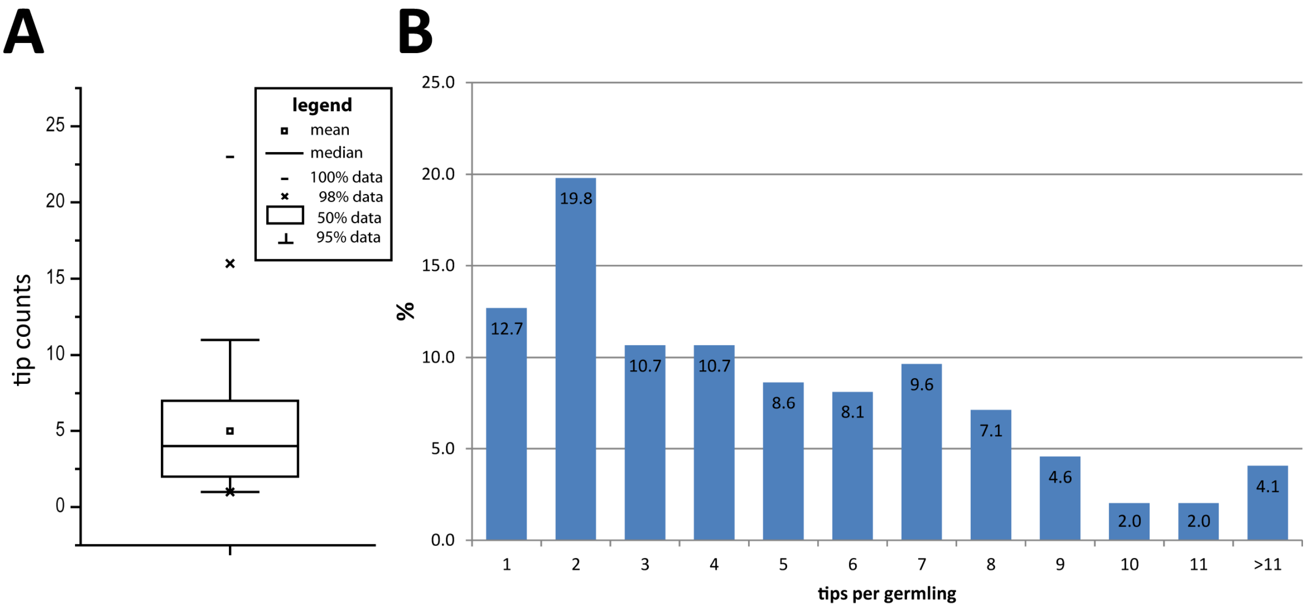


Figure4-3: Distribution of polarized growing tips in young mycelia of the 11 hours culture. **(A)** Box plot of the number of tips per observed developing young mycelium. Mean = 4,98; median = 4. **(B)** Bar diagram of the percent amount of developing young mycelia carrying different numbers of growing tips. N=192.

Based on the numbers of analyzed cells, 83% of the protein extract originated form highly polarized young mycelia and 17% from spores and germ bubbles. However, the vast majority of the biomass at this time point consists of multipolar cells, due to the massive increase in cell size. Thus, the biomass of spores and germ bubbles can most likely be neglected. This means that the protein composition determined for the cells in the 11 hours culture should closely resemble the average proteome of young mycelia. This assumption is supported by biomass measurements (Figure4-4). We incubated 5.8 mg spore biomass at 30°C. After eleven hours, the biomass had increased by a factor of six to 34.5 mg. Using the percent numbers (Figure4-1) we divided the starting spore biomass of 5.8 mg into fractions. 11% of spores represented by 0.641 mg were still present after 11 hours incubation. The 6.5% germ bubbles still observed correspond to a spore mass of 0.379 mg which increased by a factor of 1.6 to 0.606 mg at eleven hours (Figure3-9B). In total, spores and germ bubbles contributed 1.247 mg to the biomass after eleven hours. Compared to the total biomass, spores and germ bubbles only contributed 3.6% to the 34.5 mg biomass (Figure4-4A). In conclusion, 4.8 mg spores represented 82.5% of the 5.8 mg starting culture. These 82.5% increased over eleven hours by a factor of nine to 33.2 mg biomass. We applied the same calculation to the extracted amount of proteins of both time points. The protein content increased from 0.31 mg to 5.2 mg (Figure4-4B). Thus, only 1.2% of the extracted proteins at eleven hours originated from spores and germ bubbles. These data clearly demonstrated that the extracted protein composition at the 11 hours time point is characterized by cells which have efficiently established polar growth, which we termed young mycelia. Importantly, the protein concentration of the biomass increased from 5.6% to 12.2% in young mycelia (Figure4-4C).

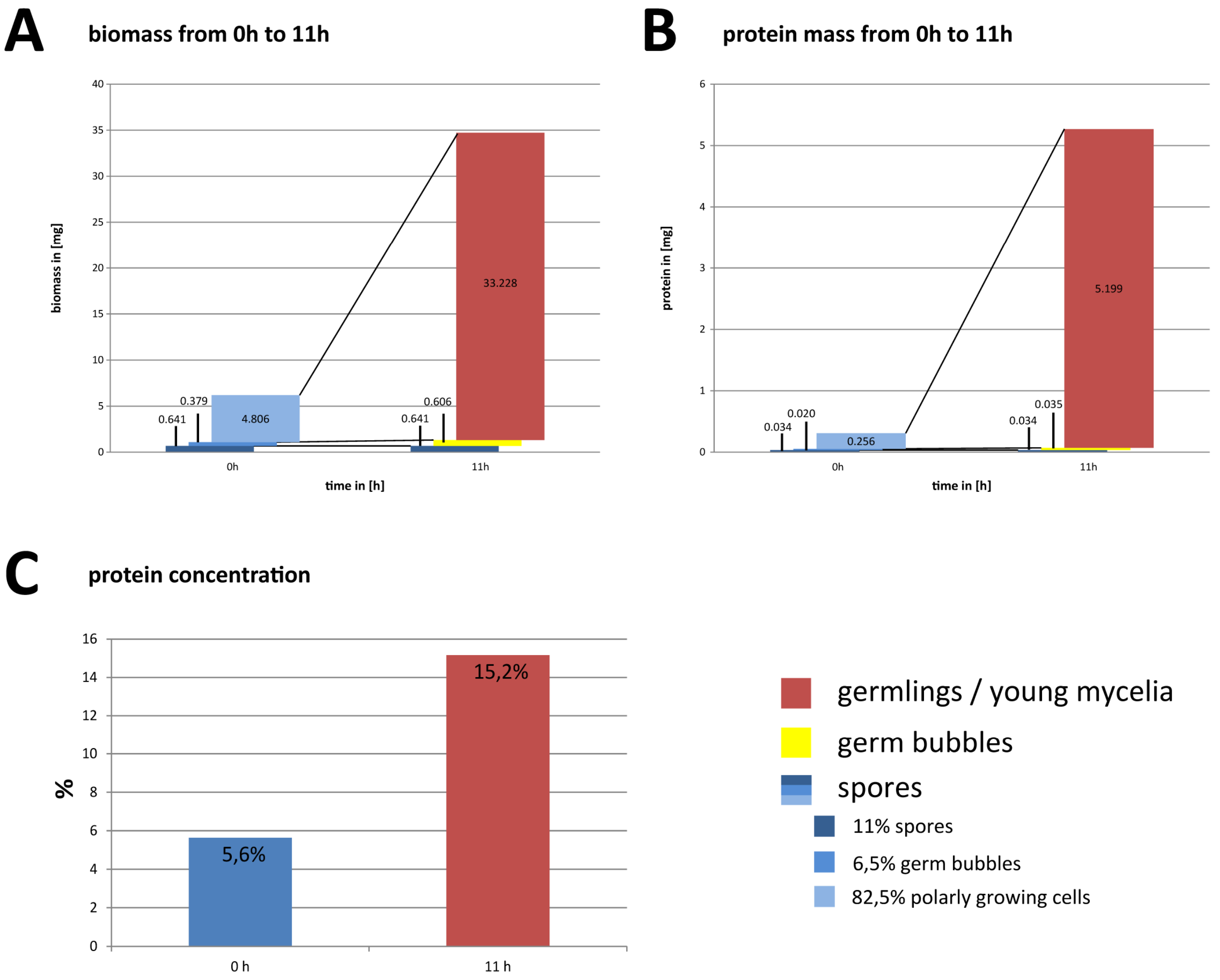


Figure4-4: Total increase of biomass and protein content during eleven hours development from spores to young mycelia. **(A)** Biomass increase differentiated according to the biomass of the observed non-developing spores (0.64 mg, dark blue), the minor biomass increase for the observed germ bubbles (0.38 mg to 0.61 mg, yellow) and the main increase for the developing young mycelia (4.8 mg to 33.2 mg, red). **(B)** Increase of the extractable protein amount similarly differentiated as in A. **(C)** Protein concentration per biomass in spores (0h) and the culture of young mycelia (11h).

Concomitantly with the acceleration of growth and the increase in protein concentration, also the number of nuclei increased. Compared to spores with one nucleus and germ bubbles with a mean of 2.5 nuclei, the number of nuclei per mycelium showed more variability. For example, a representative young mycelium of the 11 hours time point is shown in Figure4-5. The young mycelium has three original germ tubes, one with three branches and one with only one developed branch. We determined the number of nuclei after DAPI staining which visualizes double-stranded DNA by excitation at 358nm, due to the emission of blue light at 461nm (Figure4-5B). In total, 56 nuclei were counted which were approximately uniformly distributed in the cytoplasm.

In the next step, we measured the total length of the young mycelia and divided the length by the number of tips to establish a branching index (Figure4-6B). We determined an average length per tip of 17 µm (Figure4-6A). In summary, we found a huge range of tips per young mycelium (Figure4-3B) and examples of different young mycelia are shown in Figure4-6C. The unipolar germling was 19 µm long and the two bipolar germlings 29 µm and 33 µm. The two mycelia with three tips were 47 µm and 61 µm long. The mycelia with five, six, and seven tips had a total length of 76 µm, 109 µm, and 123 µm, respectively. The total hyphae length of the young mycelium with eight tips was 156 µm.

Several conclusions were drawn from these measurements. 1) The hyphal length of each young mycelium divided by the number of tips reveals the average hyphal length per polarity axis. In total, we measured 193 young mycelia and computed an average of about 17 µm per polarity axis. 2) Assuming that the first 6 µm of hyphae represent the active growth zone, approximately 35% of the surface of young mycelia are involved in fast polar surface expansion and about 65% can be viewed as hyphal surfaces with finished cell wall synthesis. 3) It is possible that only nuclei close to the growing tip express genes essential for efficient polar growth. Such polarity factors may be found at low abundances only because the proteome of the active growth zone cannot separately be determined from the proteome of the hyphae with finished cell walls. However, one can also argue that all nuclei in young mycelia have a very similar expression programme directed to form a proteome best suited to achieve fast polar growth. This view is supported by the fact, that the rare septa present in young mycelia divide the mycelium in compartments, each supporting one polar growth site. Only in older mycelia septa delineate compartments no longer involved in polar growth.

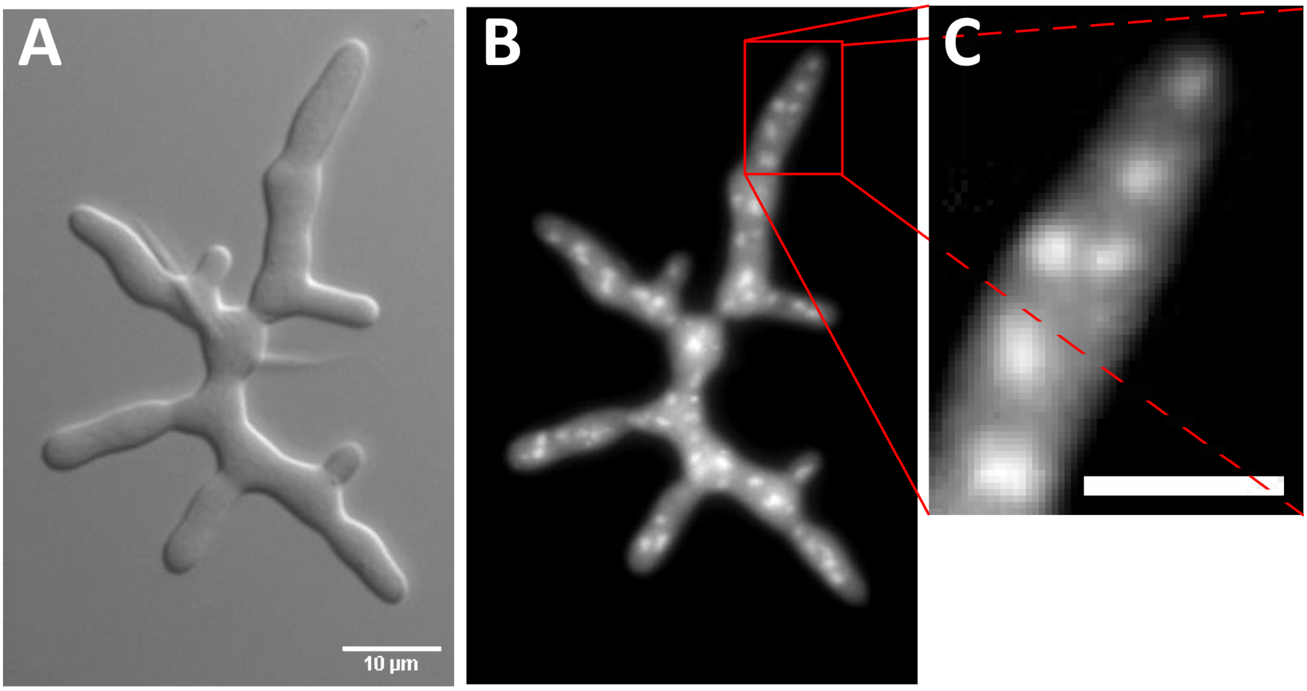


Figure4-5: Multiple nuclei in a young mycelium stained with DAPI. **(A)** DIC image of a mycelium with 8 growing tips. Scale bar 10 µm **(B)** Fluorescence image of the mycelium in A with a wavelength absorption of 358 nm and an emission at 461 nm. **(C)** Enlarged section of the hyphal tip marked in B. Scale bar 5 µm.

Characterization of a single young mycelium of the 11 hours culture

As documented in the previous section young mycelia in the 11 hours culture show a wide range of total lengths and number of polar growth sites. Originally, we only used the tip numbers to differentiate between young mycelia without taking into account the contribution of the biomass which is equivalent to the total length. Therefore, we compared the distribution of young mycelia according to the number of tips and according to the total length (Figure4-7). Based on the lengths the average young mycelium has 7.5 growing tips as calculated from the box plot distribution of Figure4-7A, and closely resembles the young mycelium shown by the image of Figure6-B. The comparison of both methods revealed a lower impact of young mycelia with fewer tips using the method weighted by length (Figure4-7B). The percent distribution of young mycelia according to tip numbers is shown in blue and according to the total length in green. With increasing length also the number of nuclei increases. For example, a young mycelium with six tips and 47 nuclei is presented in Figure4-8C. Twice as long mycelia carry about 100 nuclei (data not shown). In summary, on average one spore develops over 11 hours to a young mycelium with 7 to 8 polar growth sites and about 50 nuclei, and the amounts of protein and biomass increase by a factor of seven.

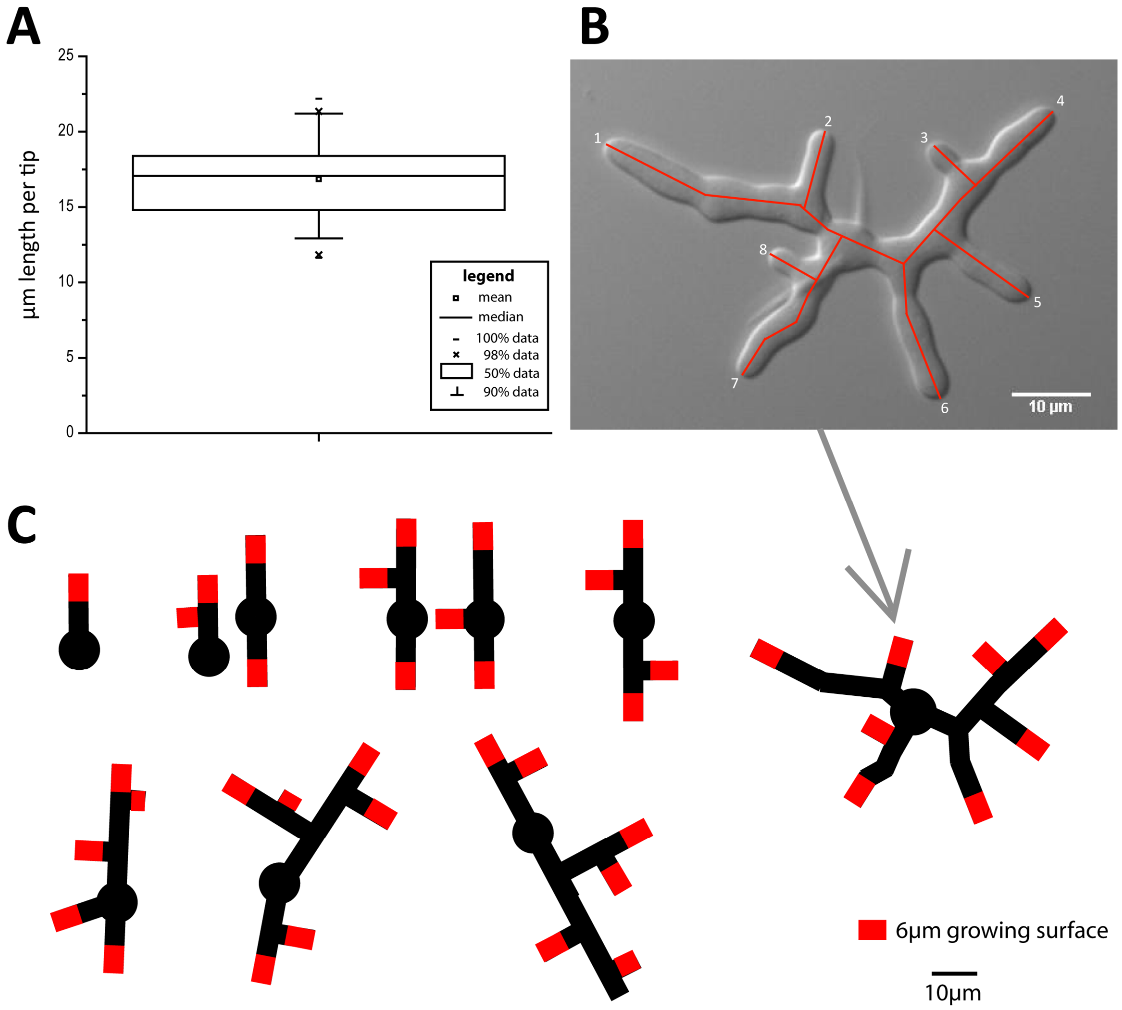


Figure4-6: Defining an average young mycelium and differentiation between growing and non-growing areas. **(A)** Box plot analysis of the total length of mycelia divided by the number of growing tips. This tip index is also called branching index. **(B)** DIC image showing the principle of the measurements for the total length of a young mycelium. Scale bar 10 µm **(C)** Schematic representation of different young mycelia observed in the 11hours culture. The tip areas with active polar growth (about 6µm) are marked in red. The black circles represent the former germ bubble.

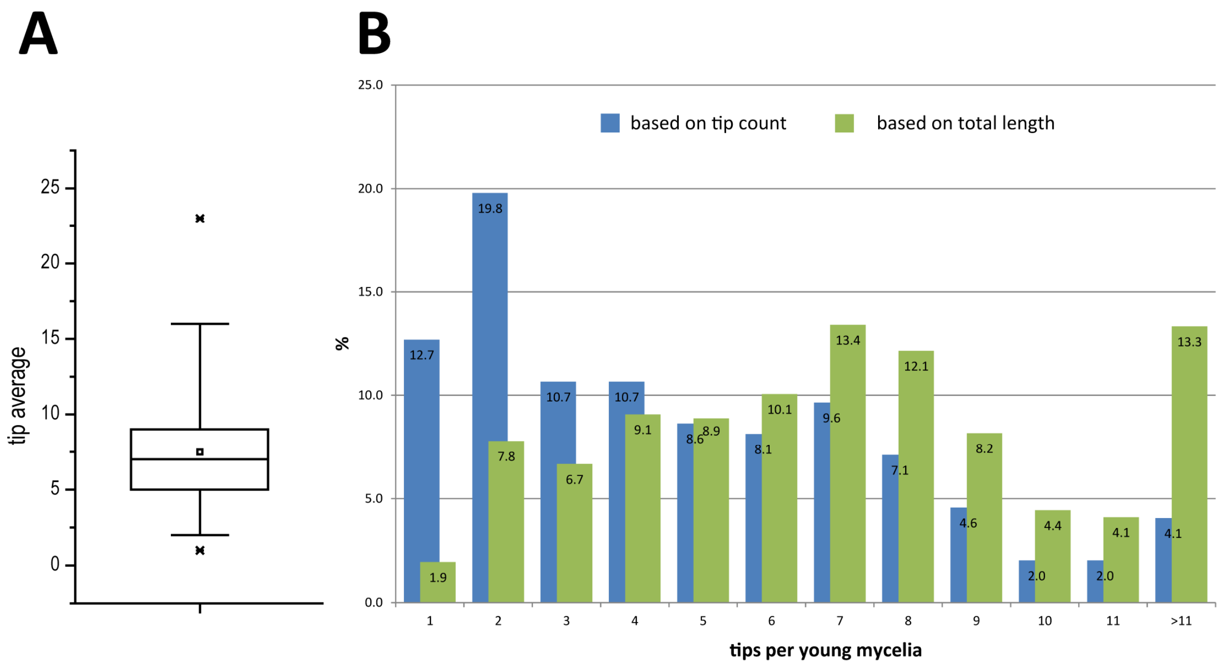


Figure4-7: Average number of tips per young mycelium according to the total length of mycelia. **(A)** Box plot of the lengths/volume distribution of young mycelia. The mean length mycelium has 7.5 tips. **(B)** Column diagram of the percent numbers of mycelia according to the number of tips (blue) or the total length/volume of all mycelia (green).

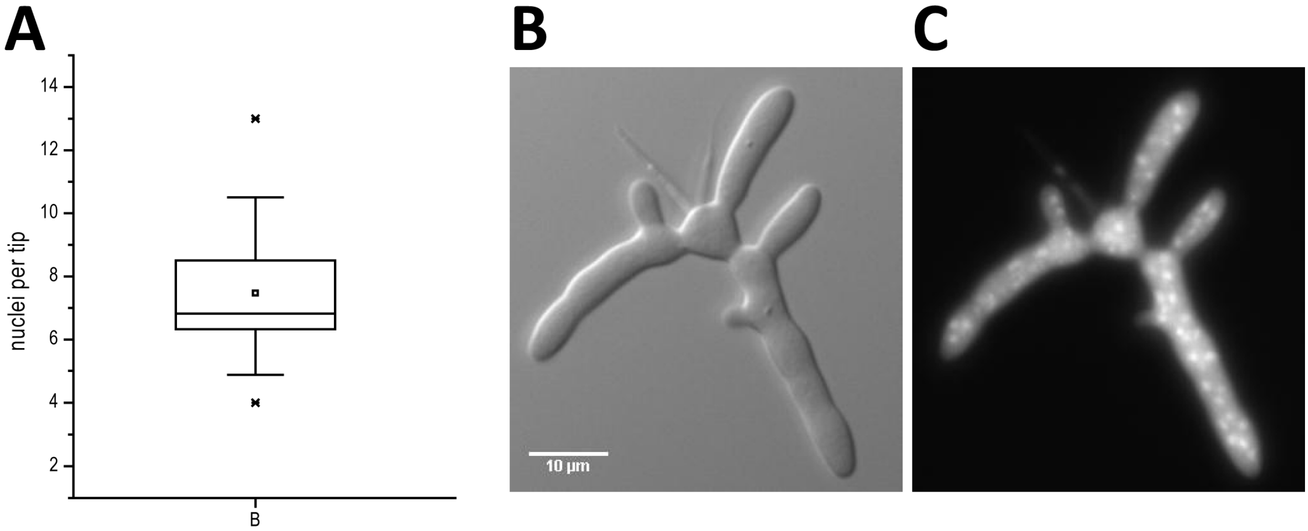


Figure4-8: Number of nuclei per tip. **(A)** Box plot of the distrubution of nuclei per tip in young mycelia. Mean = 7.5; median = 6.8 **(B)** DIC image of a young mycelium with 5 tips (polar growth axes) and 1 emerging tip (novel growth axis). **(C)** Maximum projection of a stack of 7 fluorescence images of after DAPI staining.

Although, we defined a standard young mycelium, we did not assume a uniform protein distribution and identical biological processes along the hyphae of the young mycelium. Most likely these factors differ. In nature, proteins can be localized to specific cell structures and can have gradients within cells. For example, polarized growth of *A. gossypii* requires localization of specific factors to the very hyphal tip to ensure efficient endocytosis (Köhli et al., 2008). We defined the growth zones in hyphal tips with 6 µm. This led to the assumption of two different cell compartments in young mycelia. One compartment is the area of novel cell surface synthesis, including the essential maturation of cell walls. The other compartment comprises the zones of non-growing surface with finished cell walls. We measured the length of the hyphal network of each mycelium and subtracted 6 µm for each growing tip. Next, we calculated the percentage of both compartments (Figure4-9). By using the index of the active polarized growth zone, we found that the relative length with finished cell walls slowly increases with the increase of total hyphal lengths and number of tips. The percentage of non-growing hyphal segments (with a most likely constant distribution of proteins) increases only from 60% to almost 70% in young mycelia with length. Even though, we cannot not specify the exact protein abundance in both zones, it is important to note that the compartments with finished cell walls increased during development.

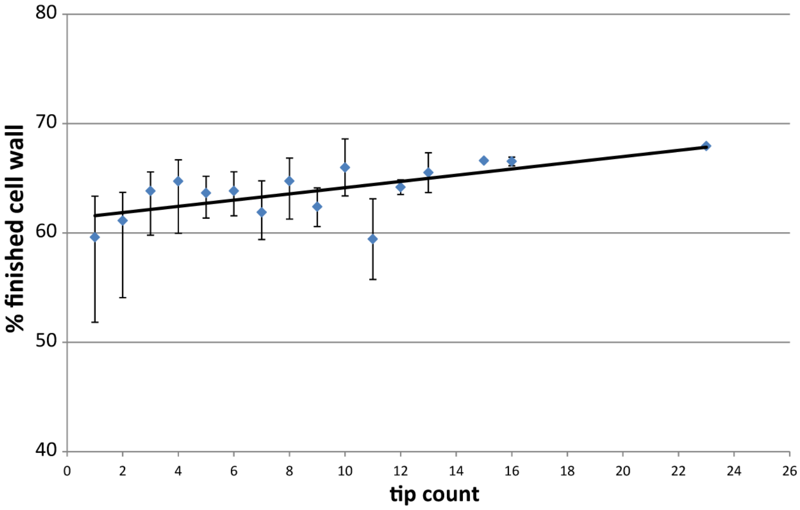


Figure4-9: Increase of non-growing hyphal surfaces (finished cell wall) to actively growing tip surfaces (new cell wall) with increasing numbers of tips. Linear fit based on the mean (green dots) of percentage of non-growing areas to the total length of the young mycelia.

Highly abundant proteins in young *A. gossypii* mycelia

The proteome analysis was performed as described in the previous two Chapters. Again, nearly 3,800 proteins (82% of all proteins predicted by the genome) could be detected and their copy numbers determined. A substantial part of these proteins could be expressed either from nuclei in the active growing tip or in hyphae area with a finished cell wall. Since, it is not known if nuclei show different expression levels depending of their positions to the tip, the contribution of each nuclei could not be distinguished.

The ranking of the 20 most abundant proteins is listed in Table4-2. Their potential functions are mentioned in Table4-3. The functional descriptions are based on the known functions of the syntenic homologs in *S. cerevisiae* copied from SGD (*Saccharomyces* Genome Database). The 20 most abundant proteins can be assigned to four functional groups:

1. Translation (Tef1, Anb1)
2. Protein assembly and aggregation (Cpr1, Ssb1)
3. Carbon metabolism (Tdh3, Fba1)
4. Ribosomal (Rps29, Rpl2, Rpl28, Rps28, Rpl38, Rps8, Rps25, Rpl27, Rps17, Rpl4, Rps0, Rpl17,Rps1, Rpl7, and Rpl8)

Compared to the spore proteome (Chapter2) and the germ bubble proteone (Chapter3), huge changes affected the top 20 abundant proteins. Interestingly, fifteen (75%) of the 20 highest expressed proteins are now ribosomal proteins of the large and small subunit. Also Anb1, a translation elongation factor, increased in abundance and change its rank from 50 to 6. The protein Ssb1 is a member of the heat shock family (Werner-Washburne et al., 1987) and is associated with ribosomes (Pfund et al., 1998). Ssb1 functions in folding newly translated polypeptide chains and prevents misfolding and aggregation of these proteins (Becker and Craig, 1994; Bukau and Horwich, 1998). Ssb1 together with ribosomal proteins was shown to regulate general protein expression (Lopez et al., 1999). This fact is also reflected in the increase of the Ssb1 protein abundance and accordingly the ranking.

Table4-1: Ranking of the 20 most abundance proteins in *A. gossypii* young mycelia. (*1) All detected peptides are taken into account and divided by all possible peptides - All-Method. (*2) Average of the top3 abundant proteins – Top3-Method.

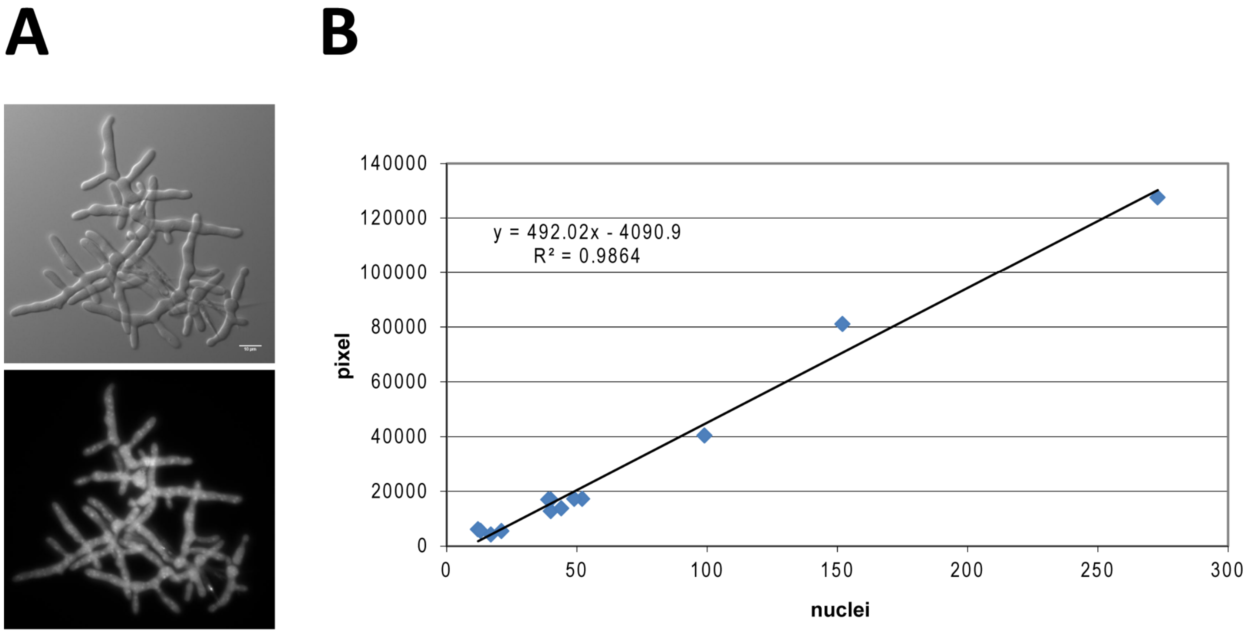
			relative abundance (all *)			relative abundance (top3 *^{2})			detected	possible	protein	protein size	
no.	accession	common name	0h	5h	11h	0h	5h	11h	peptieds	peptides	size [aa]	[MW]	Score
1	AER052W	Rps29A, Rps29B	9067	30665	125383	11810	40586	165415	5	4	56	6605	231
2	ADL370C	Tef1, Tef2	51734	59134	94159	203742	195965	356622	66	27	458	49876	3384
3	ADL127C	Rpl2A, Rpl2B	6389	14927	75232	27666	60540	344846	18	15	254	27320	939
4	AGR261W	Rps28B, Rps28A	1166	1031	42747	1368	1308	56758	8	4	67	7528	345
5	AFR501C	Rpl38	425	4379	23204	497	6465	34220	12	5	78	8822	606
6	AFR356C	Anb1, Hyp2	5265	7499	22796	12043	13184	44597	11	8	157	17141	679
7	ABR068C	Fba1	18909	16659	22545	78537	68804	93362	35	17	360	39319	2287
8	AGR375W	Rps8A, Rps8B	8272	8898	21231	28602	28389	73375	16	12	200	22485	1029
9	AER031C	Tdh3	10779	15574	20402	33108	53899	62670	43	20	331	35479	2601
10	AFR413C	Rpl27A, Rpl27B	4656	5368	19702	12906	12125	39733	18	10	136	15426	864
11	ABR033C	Rps25A, Rps25B	6216	7295	18101	16122	18826	46400	13	8	108	11980	639
12	ABL174C	Ssb2, Ssb1	3401	6458	15484	20137	33484	67539	67	36	613	66194	4369
13	AER319W	Rps17B, Rps17A	1075	3269	14986	1780	4786	22030	16	8	136	15818	906
14	AAR191C	Rpl4A, Rpl4B	3809	5160	14959	15595	19178	52385	32	20	390	42286	1972
15	ACL140C	Rps0A, Rps0B	6994	10399	14237	33504	50431	69143	14	16	253	27957	756
16	AEL137W	Rpl17A, Rpl17B	6990	8168	13736	20548	22393	33580	20	11	184	20594	1331
17	AER131C	Rps1B, Rps1A	4562	9117	12926	15520	33275	35080	46	19	256	28884	2363
18	AGL177C	Cpr1	15950	10534	12599	44902	25796	30465	17	11	162	17384	1077
19	AFL082W	Rpl7A, Rpl7B	4405	6695	12342	20035	30646	69089	25	20	243	27410	1301
20	ACL070C	Rpl8B, Rpl8A	4556	6345	12301	16993	21709	45257	34	23	305	33510	1918

Protein copies per nucleus in an average young mycelium

We determined protein copy numbers per nucleus through AQUA peptides quantified by multiple reaction monitoring (MRM). We used 40 stable isotope-labeled peptides, which were spiked into the tryptic peptide (mixture concentration of 2 mg/ml) with amounts of 2 or 20 pmol. As described above, about 99% of the isolated proteins originated from polarly growing cells. As in the previous Chapter we had to correlate the amount of isolated proteins to the number of nuclei in the cells use for protein extraction. First, we measured cross-section areas of different young mycelia and then counted the associated nuclei. We created a standard curve using the determined cross-section areas of single young mycelia and compared them to their nuclei count (Figure4-10B). Next, we measured the total area of a defined volume. In 3 µl of the serial dilution, we identified a total cross-section area of $1.68\text{x}10^7$ pixel within 65 counted sections. With the linear regression y = A + Bx, we determined $3.41\text{x}10^4$ in 3 µl. Finally, we recalculated the nuclei number with the dilution factors to the starting solution (Figure4-10C). We spiked the AQUA peptides to $2.6\text{x}10^8$ nuclei. With the coefficient of determination R-Square, we determined a correlation of 0.99. On average, one nucleus was represented by an area of 407 pixels.

Table4-2: Description of the syntenic homolog of *S. cerevisiae* for the top 20 highest ranked proteins in young mycelia according to the *Saccharomyces* Genome Database (SGD).

no.	accession	common name	name description	SGD description of the firth *S. cerevisiae* homolog
1	AER052W	Rps29A, Rps29B	Ribosomal Protein of the Small subunit	Protein component of the small (40S) ribosomal subunit
2	ADL370C	Tef1, Tef2	Translation Elongation Factor	Translational elongation factor EF-1 alpha; also encoded by TEF2; functions in the binding reaction of aminoacyl-tRNA (AA-tRNA) to ribosomes
3	ADL127C	Rpl2A, Rpl2B	Ribosomal Protein of the Large subunit	Protein component of the large (60S) ribosomal subunit
4	AGR261W	Rps28B, Rps28A	Ribosomal Protein of the Small subunit	Protein component of the small (40S) ribosomal subunit
5	AFR501C	Rpl38	Ribosomal Protein of the Large subunit	Protein component of the large (60S) ribosomal subunit
6	AFR356C	Anb1, Hyp2	ANaeroBically induced	Translation elongation factor eIF-5A, previously thought to function in translation initiation; similar to and functionally redundant with Hyp2p; undergoes an essential hypusination modification; expressed under anaerobic conditions
7	ABR068C	Fba1	Fructose 1,6-bisphosphate aldolase	Fructose 1,6-bisphosphate aldolase, required for glycolysis and gluconeogenesis; catalyzes conversion of fructose 1,6 bisphosphate to glyceraldehyde-3-P and dihydroxyacetone-P
8	AGR375W	Rps8A, Rps8B	Ribosomal Protein of the Small subunit	Protein component of the small (40S) ribosomal subunit
9	AER031C	Tdh3	Triose-phosphate DeHydrogenase	Glyceraldehyde-3-phosphate dehydrogenase, isozyme 3, involved in glycolysis and gluconeogenesis; tetramer that catalyzes the reaction of glyceraldehyde-3-phosphate to 1,3 bis-phosphoglycerate; detected in the cytoplasm and cell wall
10	AFR413C	Rpl27A, Rpl27B	Ribosomal Protein of the Large subunit	Protein component of the large (60S) ribosomal subunit
11	ABR033C	Rps25A, Rps25B	Ribosomal Protein of the Small subunit	Protein component of the small (40S) ribosomal subunit
12	ABL174C	Ssb2, Ssb1	Stress-Seventy subfamily B	Cytoplasmic ATPase that is a ribosome-associated molecular chaperone, functions with J-protein partner Zuo1p; may be involved in the folding of newly-synthesized polypeptide chains; member of the HSP70 family
13	AER319W	Rps17B, Rps17A	Ribosomal Protein of the Small subunit	Protein component of the small (40S) ribosomal subunit
14	AAR191C	Rpl4A, Rpl4B	Ribosomal Protein of the Large subunit	Protein component of the large (60S) ribosomal subunit
15	ACL140C	Rps0A, Rps0B	Ribosomal Protein of the Small subunit	Protein component of the small (40S) ribosomal subunit
16	AEL137W	Rpl17A, Rpl17B	Ribosomal Protein of the Large subunit	Protein component of the large (60S) ribosomal subunit
17	AER131C	Rps1B, Rps1A	Ribosomal Protein of the Small subunit	Protein component of the small (40S) ribosomal subunit
18	AGL177C	Cpr1	Cyclosporin A-sensitive Proline Rotamase	Cytoplasmic peptidyl-prolyl cis-trans isomerase (cyclophilin), catalyzes the cis-trans isomerization of peptide bonds N-terminal to proline residues
19	AFL082W	Rpl7A, Rpl7B	Ribosomal Protein of the Large subunit	Protein component of the large (60S) ribosomal subunit
20	ACL070C	Rpl8B, Rpl8A	Ribosomal Protein of the Large subunit	Protein component of the large (60S) ribosomal subunit



C

	calculation	result
total area		$1,68*10^7$
nuclei in 3µl	$x = (1,68*10^7 - 4090,9) / 492,02$	$3,41*10^4$
in 200µl DAPI	$= 3,41*10^4 / 3 * 200$	$2,27*10^6$
1/4 used from fix used for DAPI staining	$= 2,27*10^6 * 4$	$9,10*10^6$
1ml used from 150ml	$= 9,10*10^6 * 150$	$1,36*10^9$
protein concentration (5,25mg/ml)	$= 1,36*10^9 / 2,25$	$2,60*10^8$
total amount used 2mg	$= 2,60*10^8 * 2$	$5,20*10^8$
		$5,20*10^8$

Figure4-10: Total number of nuclei in mycelia needed to extract 2 mg of protein. **(A)** DIC and fluorescent image of an DAPI stained young mycelium. Scale bar 10 µm. **(B)** The areas of thirteen young hyphae were measured and the nuclei were counted in these areas to generate a regression line to relate number of pixels to number of nuclei. R^2 is 0,99. **(C)** Calculations to gain the total number of nuclei in the mycelial biomass used for protein isolation.

To determine the copy numbers of protein per nucleus, we applied the same methods described in detail in Chapter2 and 3. First, we compared the 40 AQUA peptides with their identified endogenous partners (Figure4-11). We were able to identify 13 of the 20 with 2 pmol spiked AQUA peptides. The majority of the detected stable-isotope labeled peptides had abundances between 1,000 and 100,000. Of the 20 pmol spiked AQUA peptides, we detected 17 MS1 spectra, ranging in abundance from about 100,000 to 1,000,000. For peptides with a spiked concentration, we found for the 2 pmol an abundance ratio of 100 and for the 20 pmol a ratio of only 10.

Table4-4 shows the determined endogenous peptide concentrations and their corresponding proteins. We determined the copy number per nucleus of 30 proteins by implicating the defined nuclei count (Figure4-10). Specifically, we used the heavy AQUA peptides and computed the concentration of the light peptides, which equals the exact concentration of the corresponding protein. For more detail of the approach please see Chapter2C. The protein copy numbers per nucleus of young mycelia ranged from eleven to about 56,000 copies. The membrane bound guanine nucleotide-exchange factor (GEF) Cdc25 (Jones et al., 1991) had the lowest copy number per nuclei with 11 proteins. In contrast, the cytoskeleton protein actin, which plays an essential role in cell polarization and endocytosis (Pruyne and Bretscher, 2000), was found with 56,676 copies per nucleus.

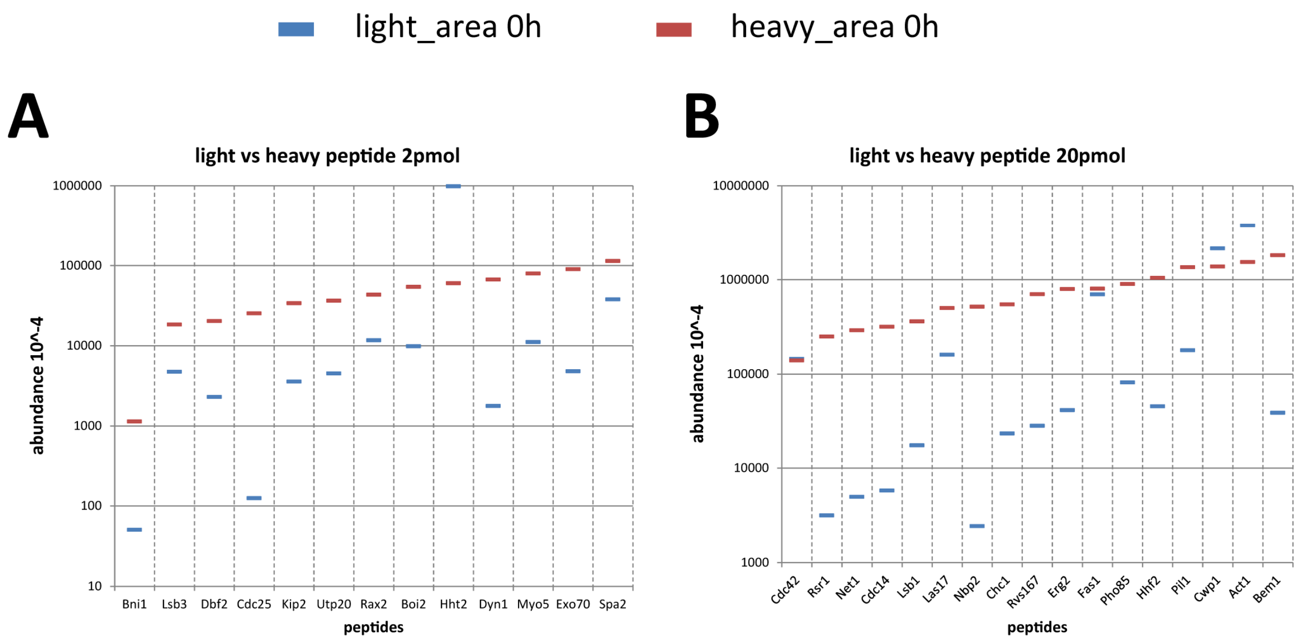


Figure4-11: Abundances of 30 pairs of heavy/light peptides in MS1-spectra for the determination of absolute quantities of reference proteins. **(A)** Red bars: Abundances of 13 AQUA peptides added as 2 pmol to the peptide mixture. Blue bars: Abundances of the corresponding light peptides. **(B)** Red bars: Abundances of 17 AQUA peptides added as 20 pmol to the peptide mixture. Blue bars: Abundances of the corresponding light peptides. The proteins from which the reference peptides were selected are indicated at the bottom of the graphs.

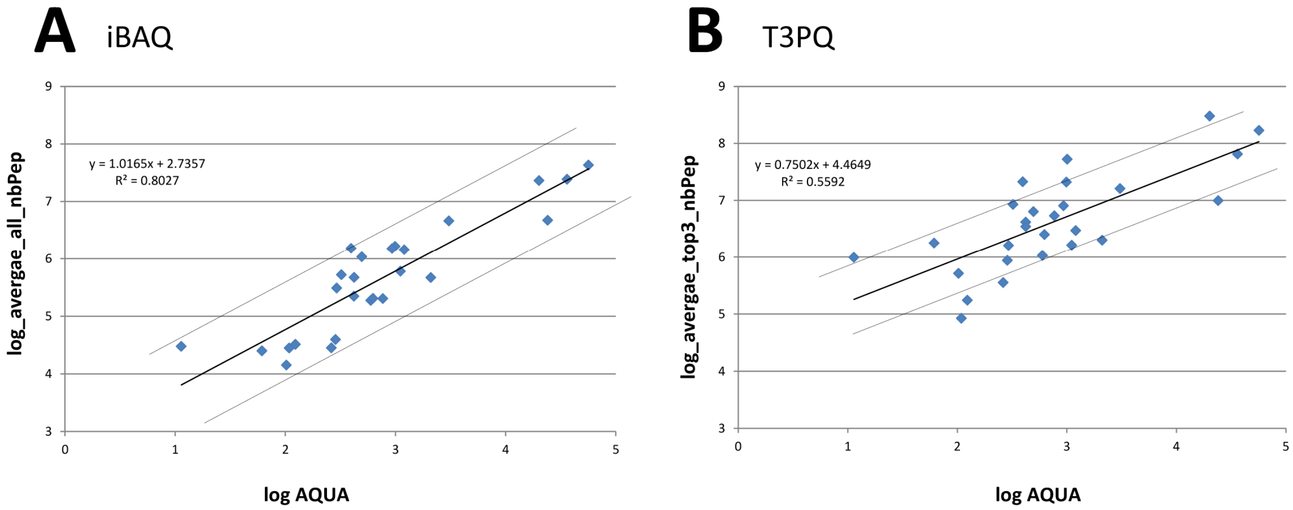


Figure4-12: Regression line of protein abundances in relation to the measured copies per nucleus. The x-axis represents the logarithm of the copy numbers calculated *via* the ratio of heavy – light peptides. The y-axis represents the logarithm of the protein abundance defined by **(A)** iBAQ and by **(B)** T3PQ. **(A)** The function values a has an error of 0.30836 and b of 0.10287. **(B)** The function values a has an error of 0.39933 and b of 0.13321.

Table4-4: Protein copies per nucleus of the 35 reference proteins based on the ratios of light to heavy peptides.

accession	common name	relativ abundance		ratio from area	heavy peptide amount in[pm]	concentration in [pm]	nuclei count; 5h post inoculation	protein concentration in pmol per nucleus	**copies per nucleus**	protein abundance / possible peptides [10^-4]	protein abundance / top3 peptides [10^-4]	log copies per nucleus	log protein abundance / possible peptides	log protein abundance / top3
		light area	heavy area											
ABR222W	Act1	37700000000	15410000000	2.4465	20	48.9292667	519891088	9.41E-08	**56676**	4296	16905	4.753	7.633	8.228
ADL202C	Hht2	9882000000	602600000	16.3989	2	32.79787588	519891088	6.31E-08	**37990**	221	348	*4.580*	*6.344*	*6.542*
ABR028C	Cwp1	21581256000	13822332800	1.5613	20	31.22664794	519891088	6.01E-08	**36170**	2437	6483	4.558	7.387	7.812
AGL093W	Cdc42	1452852000	1397068000	1.0399	20	20.79858676	519891088	4.00E-08	**24091**	470	995	4.382	6.672	6.998
AER085C	Fas1	7000000000	8060000000	0.8685	20	17.36972704	519891088	3.34E-08	**20120**	2301	30148	4.304	7.362	8.479
AGR285W	Las17	1607260000	5026000000	0.3198	20	6.39578194	519891088	1.23E-08	**7408**	8	63	*3.870*	*4.924*	*5.801*
AEL329W	Pil1	1791682000	13582920000	0.1319	20	2.63813966	519891088	5.07E-09	**3056**	460	1612	3.485	6.663	7.207
AGL242C	Pho85	816890000	9021600000	0.0905	20	1.8109648	519891088	3.48E-09	**2098**	47	201	3.322	5.671	6.302
AFL181C	Erg2	415200000	7980660000	0.0520	20	1.04051544	519891088	2.00E-09	**1205**	144	296	3.081	6.158	6.471
AFR320W	Lsb1	173820000	3628000000	0.0479	20	0.9582139	519891088	1.84E-09	**1110**	60	163	3.045	5.781	6.213
ADL201W	Hhf2	456610000	10501000000	0.0435	20	0.8696505	519891088	1.67E-09	**1007**	2710	5276	3.003	*7.433*	7.722
AER359W	Chc1	235000000	5492000000	0.0428	20	0.85579024	519891088	1.65E-09	**991**	165	2091	2.996	6.218	7.320
AFR140C	Rvs167	283406600	7044820000	0.0402	20	0.80458152	519891088	1.55E-09	**932**	151	808	2.969	6.178	6.907
ADL022C	Spa2	379054000	1143312000	0.3315	2	0.663080594	519891088	1.28E-09	**768**	20	539	2.885	5.307	6.732
AGR095W	Rax2	117200000	434000000	0.2700	2	0.540092166	519891088	1.04E-09	**626**	20	252	2.796	5.308	6.401
AEL017W	Lsb3	47500000	184000000	0.2582	2	0.516304348	519891088	9.93E-10	**598**	19	107	2.777	5.274	6.028
AEL241W	Bem1	389174000	18227000000	0.0214	20	0.42703022	519891088	8.21E-10	**495**	108	637	2.694	6.033	6.804
AEL025W	Cdc14	57892000	3182400000	0.0182	20	0.36382604	519891088	7.00E-10	**421**	47	416	2.625	5.675	6.619
AGL293C	Boi2	98848000	544202000	0.1816	2	0.363276872	519891088	6.99E-10	**421**	22	347	2.624	5.346	6.541
AAL181C	Net1	49762800	2914140000	0.0171	20	0.34152648	519891088	6.57E-10	**396**	152	2126	2.597	6.183	7.328
AEL306C	Myo5	111190000	797800000	0.1394	2	0.27874154	519891088	5.36E-10	**323**	53	853	2.509	5.720	6.931
AFR464W	Rsr1	31637000	2496900000	0.0127	20	0.25341022	519891088	4.87E-10	**294**	31	161	2.468	5.488	6.206
AEL006W	Utp20	45200000	366000000	0.1235	2	0.246994536	519891088	4.75E-10	**286**	4	88	2.457	4.596	5.942
ADR033W	Dbf2	23034000	203600000	0.1131	2	0.22626719	519891088	4.35E-10	**262**	3	36	2.418	4.451	5.553
ACR145W	Kip2	35800000	340000000	0.1053	2	0.210588236	519891088	4.05E-10	**244**	0	2	*2.387*	*2.726*	*4.251*
AFR100W	Exo70	48300000	904200000	0.0534	2	0.106834772	519891088	2.05E-10	**124**	3	17	2.093	4.511	5.243
AGL169C	Nbp2	24360000	5186000000	0.0047	20	0.09394524	519891088	1.81E-10	**109**	3	8	2.037	4.449	4.926
AFR669W	Bni1	504000	11380000	0.0443	2	0.08857645	519891088	1.70E-10	**103**	1	52	2.011	4.153	5.713
ACR258W	Dyn1	17840000	672000000	0.0265	2	0.053095238	519891088	1.02E-10	**62**	3	178	1.789	4.400	6.251
ADL038W	Cdc25	1247000	253800000	0.0049	2	0.009826636	519891088	1.89E-11	**11**	3	99	1.056	4.476	5.994
ABR008C	Hse1	-	-	-	2	-	519891088	-	**-**	38	285	*-*	*5.581*	*6.455*
AEL090C	Inp54	-	-	-	2	-	519891088	-	**-**	3	17	*-*	*4.410*	*5.243*
AFR441C	Tub1	-	-	-	20	-	519891088	-	**-**	1742	6066	*-*	*7.241*	*7.783*
AGL364C	Bnr1	-	-	-	2	-	519891088	-	**-**	0	2	*-*	*2.644*	*4.223*
AGR043W	Num1	-	-	-	2	-	519891088	-	**-**	7	296	*-*	*4.831*	*6.471*

In the following, we were able to determine the copy number for all detected proteins in young mycelia. We calculated the copy numbers via linear regression (Figure4-12) with the given formula y = a + bx. To build the regression line, we used the AQUA-peptides based protein numbers (Table4-4 column 13) versus the analogous protein abundance determined of both, All- and Top3-method (column 14 and 15). The logarithmic values were plotted in Figure4-12 and both methods iBAQ (A) and T3PQ (B) were compared. The correlation is represented by R^2 which was determined with 0.8 for iBAQ and 0.6 for T3PQ. The R-Square for the iBAQ method indicates the better correlation which should result in more accurate protein copy numbers. In addition, we validated the results of the linear regression with the copy numbers based on heavy to light peptide ratios. In Table4-6, the ratio between both results were listed for the iBAQ and T3PQ method. The ratio difference indicated if the copy numbers were accurately determined with the linear regression. We compared the mean ratios of iBAQ and T3PQ to gain a better understanding of the results of the two methods. Concluding, the iBAQ method showed a better correlation using isotope-labeled peptides expressed by the higher R-Square and low ratio of the copy number determination. For the iBAQ method, we identified most of protein copy numbers in a close ratio. Only for the two histone proteins Hht2 and Hhf2 huge difference were determined. Both proteins also differed in the copy number to each other by a factor over ten. Furthermore, histones H3 and H4 interact as a heterodimer. Two of these H3-H4 heterodimers together with two heterodimers of histone H2A and H2B form the histone core octamer called nucleosome (Schafer et al., 2005). Thus, Hht2 and Hhf2 should be within the same protein copy number rage. In spores, we found both proteins with same copy numbers (Table2-19). This changed from spores to young mycelia indicating alterations in the peptide detection. The decreased detection of the Hhf2 peptide most likely resulted from modifications in the endogenous protein during digestion. This is supported by the fact that the determined heavy peptides ratio between Hht2 and Hhf2 in spores of ten (Table1-18) is compared to young mycelia (Table4-4) stable with a ratio of 17.

Discussion

Within eleven hours, an *A. gossypii* spore develops to a young mycelium. On average, a young mycelium has seven hyphae tips, a total hyphal length of about 120 µm, and approximately 50 nuclei. Spores and germ bubbles, present in the eleven hour cell mixture, did not affect the entire protein content. Young mycelia have a huge increased size compared to the relatively small volume of spores and germ bubbles. At most, only 1% of the total extracted proteins originate from remaining spores and germ bubbles. At the eleven hour time point, polar cells morphologically ranges from a single germ tube to a young mycelium with over 20 established polarity axes. In young mycelia, morphological differences as well as molecular and intracellular alterations can be identified. For example, the plasma membrane and the cell wall at growing tips is constantly remodeled by adding lipids and proteins. In contrast, it is suggested that older and non-expanding parts of the mycelia are surrounded by a completed membrane and cell wall. Therefore, we defined the first 6 µm of the hyphae tips as active and growing (Köhli et al., 2008) and the rest as finished mycelium.

We defined the general behavior of proteins based on their abundance changes from spores to young mycelia. Proteins which were only highly expressed at certain locations led to conflicts and problems in defining accurate changes of protein abundances. In general, proteins that strongly increased their abundance are most likely essential for polar growth and for building strong hyphal cell walls. In contrast, proteins decreased in abundance are actively degraded and are therefore effecting sporulation or germination. Protein abundances which are stable or minor decreased are actively expressed, due to the increased protein concentration per biomass from 5% to 15% in young mycelia. In addition, the protein amount of a young mycelium is about 20 times higher than in spores (Figure4-4). This is represented by the overall expression of ribosomal proteins, which are the most abundant protein group in young mycelia. Because of the higher ribosomal count, the biomass has increased faster. The protein concentration in young mycelia reaches over 15%. It is not in detail understood why the protein content increased disproportionately to the biomass. Most likely the high content of wall polysaccharides in spores is the reason for the relatively low protein content. Importantly, we compared a single mono-nucleated cell with a developed multi nucleated cell. By comparing both cell stages, the major volume differences had to be considered. The volume of an average young mycelium is calculated to be 1300 µm^3, which is about 35 times larger than the spore volume of 35 µm^3 (Figure2-1). Compared with the twentyfold biomass difference between young mycelia and spores, these data strongly indicates an enormous surface expansion over time.

Table4-6: Evaluation of the protein copy numbers defined *via* the AQUA peptides and the calculated protein copies per spore using the regression line. *1) Intensities of all detected peptides divided by the number of possible peptides for a protein – iBAQ. *2) The average of the 3 peptides with the highest abundance was used to calculate the copies per nucleus – T3PQ. *3) Mean was computed without outliers that are indicated light gray.

accession	common name	copies per nucleus	cpn all *1	cpn top3 *2	ratio all *1	ratio top3 *2
ABR222W	Act1	**56676**	65746	103786	-1.16	-1.83
ADL202C	Hht2	**37990**	3547	587	10.71	64.68
ABR028C	Cwp1	**36170**	37638	28927	-1.04	1.25
AGL093W	Cdc42	**24091**	7459	2377	3.23	10.13
AER085C	Fas1	**20120**	35580	224413	-1.77	-11.15
AGR285W	Las17	**7408**	142	60	*52.09*	*122.76*
AEL329W	Pil1	**3056**	7300	4526	-2.39	-1.48
AGL242C	Pho85	**2098**	772	281	2.72	7.45
AFL181C	Erg2	**1205**	2326	472	-1.93	2.55
AFR320W	Lsb1	**1110**	990	214	1.12	5.18
ADL201W	Hhf2	**1007**	41778	21980	-41.49	-21.83
AER359W	Chc1	**991**	2666	6403	-2.69	-6.46
AFR140C	Rvs167	**932**	2435	1802	-2.61	-1.93
ADL022C	Spa2	**768**	338	1051	2.27	-1.37
AGR095W	Rax2	**626**	339	381	1.85	1.64
AEL017W	Lsb3	**598**	314	121	1.90	4.93
AEL241W	Bem1	**495**	1754	1314	-3.54	-2.65
AEL025W	Cdc14	**421**	778	744	-1.85	-1.77
AGL293C	Boi2	**421**	370	585	1.14	-1.39
AAL181C	Net1	**396**	2463	6544	-6.22	-16.52
AEL306C	Myo5	**323**	864	1937	-2.67	-6.00
AFR464W	Rsr1	**294**	511	209	-1.74	1.40
AEL006W	Utp20	**286**	68	93	4.23	3.07
ADR033W	Dbf2	**262**	49	28	5.39	9.28
ACR145W	Kip2	**244**	1	1	*249.20*	*469.82*
AFR100W	Exo70	**124**	56	11	2.22	11.40
AGL169C	Nbp2	**109**	48	4	2.25	26.49
AFR669W	Bni1	**103**	25	46	4.16	2.23
ACR258W	Dyn1	**62**	43	240	1.43	-3.87
ADL038W	Cdc25	**11**	52	109	-4.69	-9.93
ABR008C	Hse1	**-**	630	449	-	-
AEL090C	Inp54	**-**	44	11	-	-
AFR441C	Tub1	**-**	27052	26474	-	-
AGL364C	Bnr1	**-**	1	0.5	-	-
AGR043W	Num1	**-**	115	473	-	-
				mean	1.82	5.60

Chapter 5: Developmental proteomics from spores to young mycelia (a preliminary view)

Chapter 5: Developmental proteomics from spores to young mycelia (a preliminary view)

Introduction

The life cycle of the filamentous fungus *A. gossypii* is relatively simple. This and the smallest genome of a free-living eukaryote (Dietrich et al., 2004) make it a perfect target to gain understanding of protein profile changes during the life cycle. The different stages of the life cycle and the corresponding distribution of the cell types are presented in Figure5-1. A major difficulty is the asynchrony of the development. Except for the spore stage each stage consists of a mixture of all types. Until the third hour after inoculation no morphological changes could be observed. After five hours at standard growing conditions 35% of the spores formed a germ bubble of different size and the rest of the spores kept the initial shape. At seven hours only 30% spores were detected and 70% consisted of a mixed population with 20% germ bubbles, 30% unipolar germlings and ca. 20% bipolar germlings some showing already a branch. Nine hours after inoculation of spores the population of cells contained 20% spores, 11% germ bubbles, 18% unipolar germlings and ca. 50% young mycelia and bipolar germlings. At eleven hours we observed over 70% fast growing young mycelia and about 10% each of spores and germ bubbles and germlings with germ tubes. The high resolution proteome analysis was performed with the zero, five and eleven hour populations.

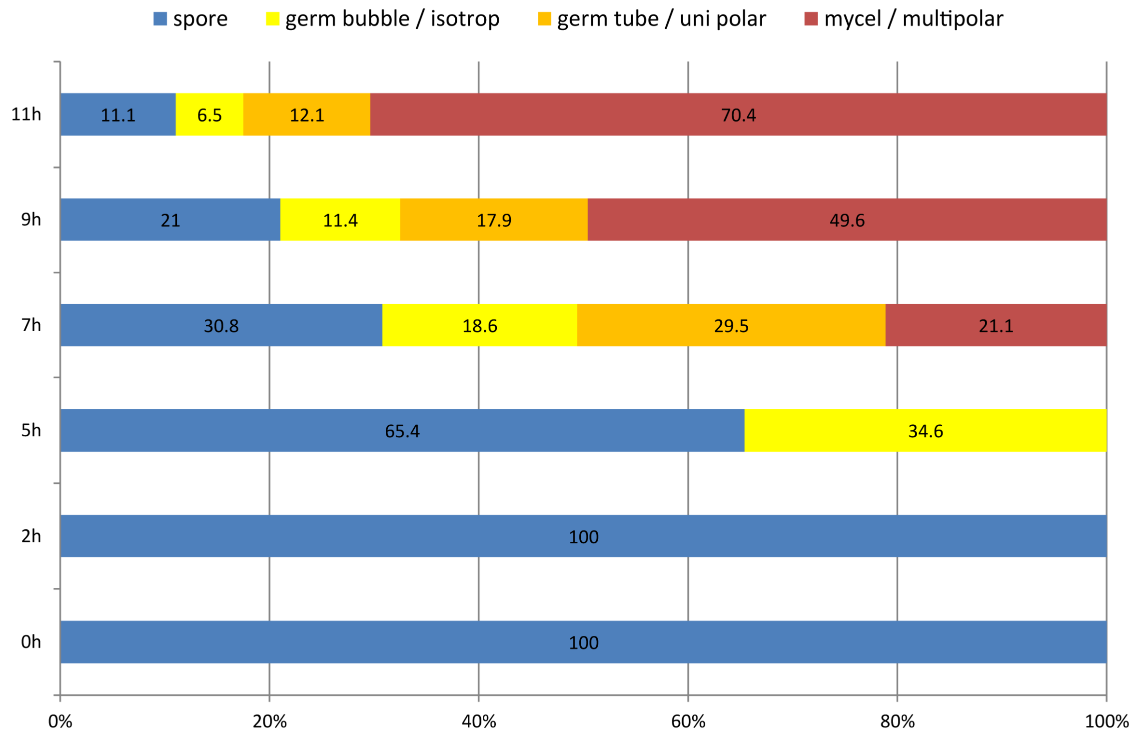
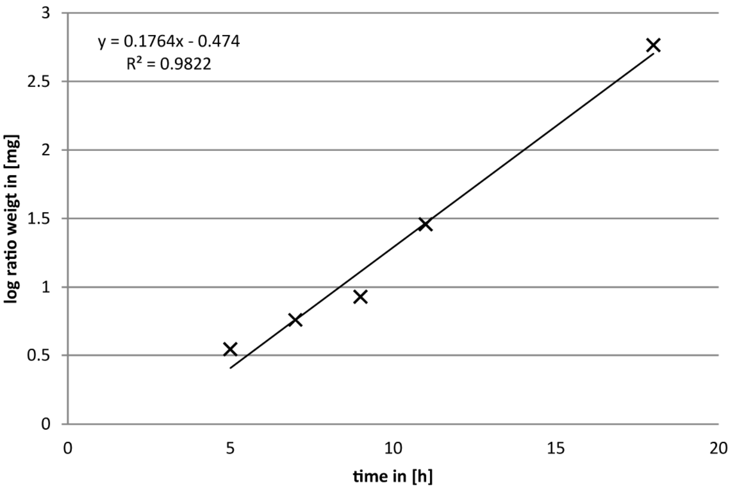


Figure5-1: Cell type composition in relation to the inoculation time of the spores in liquid medium. Cells are differentiated based on the morphological shape in spores, germ bubbles, unipolar germlings, and young mycelia including bipolar germlings, n=2,453.



	0h-5h		0h-7h		0h-9h		0h-11h		0h-18h	
time in [h]	0	5	0	7	0	9	0	11	0	18
improved weight in [mg]	2.17	5.67	4.27	10.00	4.71	13.21	5.57	34.22	5.57	586.11
mass gain in [mg]		3.50		5.72		8.50		28.65		580.54
log ratio		0.54		0.76		0.93		1.46		2.76

Figure5-2: Biomass increase from zero to eighteen hours after inoculation of spores. Subtraction of the spore weight, based on the cell population for each time point to improve the data.

During the first 18 hours of after inoculation of spores, we could observe a logarithmic increase of the biomass. Figure5-2 shows the Log scale of the biomass in mg per time plotted on the ordinate. From zero to five hours the biomass 2.17 mg spores (35% of the inoculated spores) increased 2.6 times to 5.67 mg biomass of germ bubbles. From zero to eleven hours the biomass of 5.6 mg spores (80% of the inoculated spores) increased to 586.1mg biomass of young mycelia, 100 times more than the starting biomass. This enormous increase in biomass allowed us to exclude the remaining traces of spores and germ bubbles in the characterization of the young mycelia proteome (see Figure5-1). The equation based on the linear fit results in a high coefficient of determination R^2 with 0.98. By solving the equation for x we calculated a pseudo generation time of 1.7h. Because of the asynchronous mitoses in *A. gossypii* and the variation of the general nuclear duplication time, thus generation time could not be exactly defined. But with the duplication of the biomass we could estimate the “pseudo generation” time with approximately 100 minutes, which is similar to *S. cerevisiae* growing in full medium of the same temperature. During the development from a spore to young mycelia the protein concentration with respect to biomass increase shows a remarkable increase. Spores have an average of 5.6% protein of the total biomass which increase during growth to 9.2% at seven hours up to a maximum of about 15% in young mycelium (Figure5-3). Therefore, a smaller cell volume is needed to extract the same amount of proteins. The lower protein concentration in the spore stage most likely reflects high concentration of sugar polymers in the spore wall needed to protect the spore. The increase of the protein concentration over time is not completely understood but the thick protein layer in the outer hyphal wall may be one of the reasons for this increase.

This chapter was written as a preliminary attempt to use the data sets for getting a first view of developmental proteomics in *A. gossypii* and should therefore not be seen as a final result. It is more a stimulus for the continuation of this developmental proteomics project as a postdoc aiming at high resolution and quantitative proteomics of three additional stages of the *A. gossypii* life cycle. Together with the three stages described in this PhD-thesis a complete view of developmental proteomics of *A. gossypii* should than be possible.

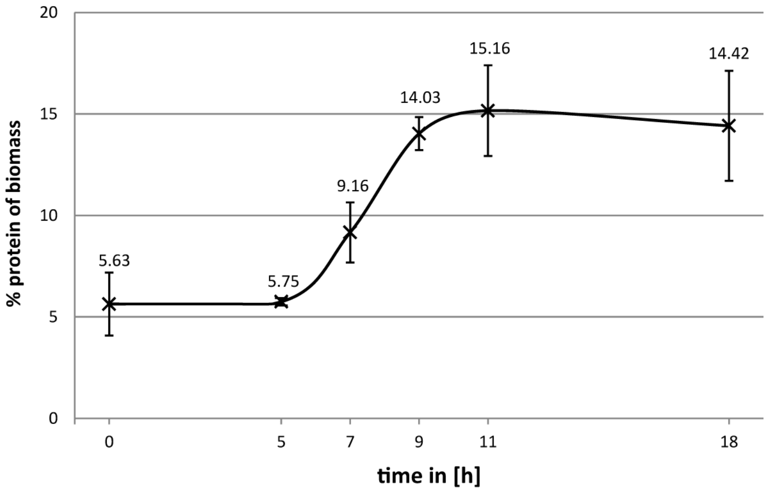


Figure5-3: Protein concentrations of the cell population stages from 0 to 18 hours. Spores were inoculated in AFM and uniform distributed by vortexing. Spore solution was homogeneous separated into cultures for six time points (0, 5, 7, 9, 11, and 18 hours) each with biological replicas. The cell weights were measured via a microbalance after dehydration of the samples using a SpeedVac. The cell samples were lysed and the protein concentrations were measured via Bradford protein assay. The values of the biomass and protein mass were computed results in the protein concentrations of the different *A. gossypii* cell stages.

Results

To gain an improved overview about the increase / decrease of all 3,800 proteins we simultaneously looked at all three time points and clustered proteins in groups according to their abundance changes during the elven hours of after inoculation of spores. The abundances were computed into the logarithm to the base 10 and the five and eleven hour stages were divided by the spore stage, which functions as a starting point of the protein characterization. Correlation between the different time points, by transforming the protein abundances into logarithm to the base 2, shows a high variability. In Figure5-4 the log2 abundance values of two different time points are plotted on the y- and x-axis. The pairs plots in Figure5-4A shows the log2 of the five hour abundance on the ordinate and the values for the zero hour time point on the abscissa. The eleven and zero hour time points are compared in B. In Figure5-4C the abundance of the five hour values are presented at the x-axis and the elven hour values are plotted at the y-axis. The Figure5-4 shows an increase in the variance from A over C to B. This indicates the more the time points are apart the higher the distribution of the data points. The five hour time point is closer to the zero hour time point than to the eleven hours. The zero hour and eleven hour time points display huge differences in the abundance of proteins, which indicates strong alterations in the protein expression or stability. These analyses clearly represent the dynamics of the *A. gossypii* proteome during the development.

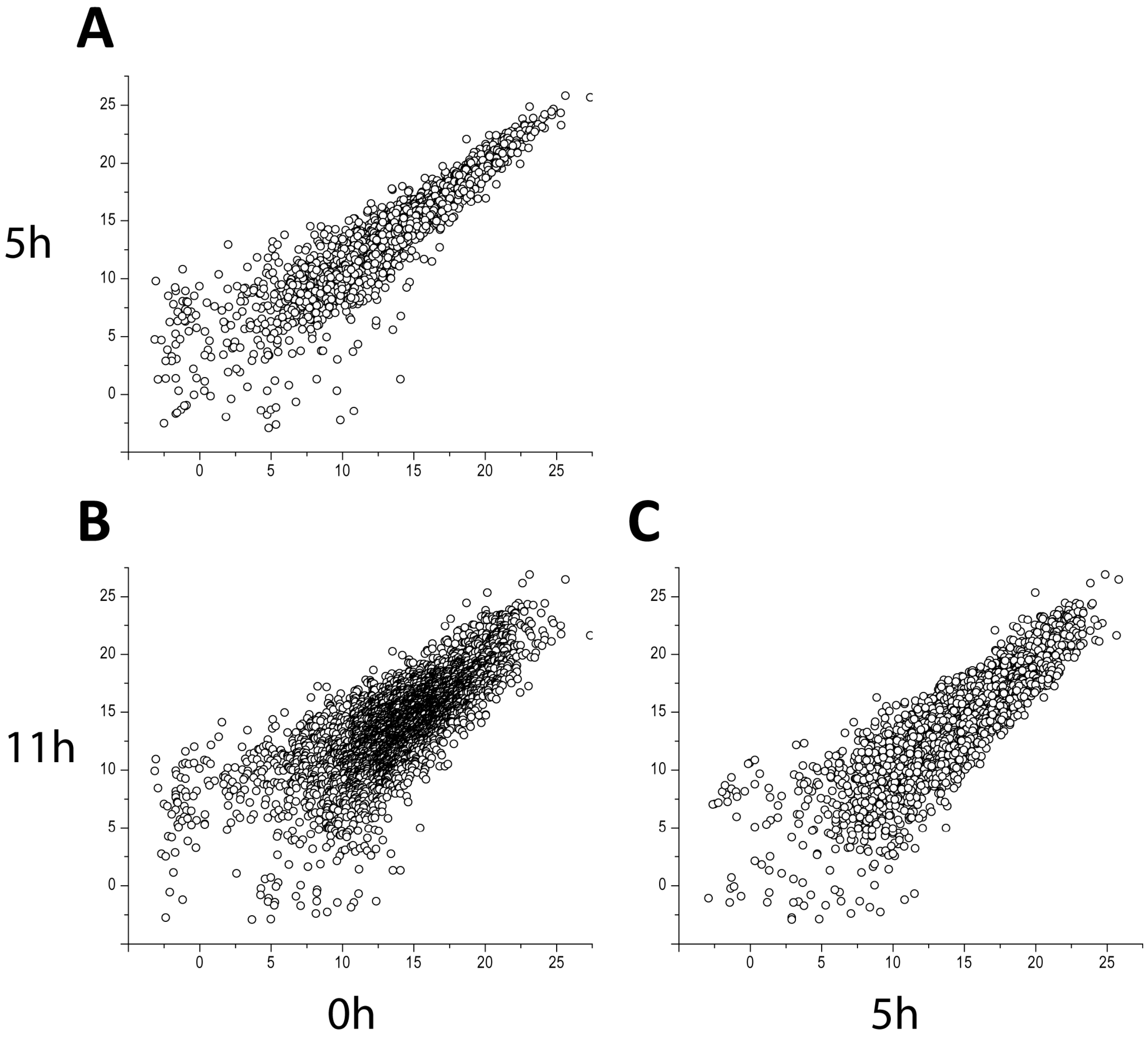


Figure5-4: Pairs plots of log2 transformed protein abundances. Correlation between: five and zero hours **(A)**, eleven and zero hours **(B)**, and eleven and five hour **(C)**.

Protein abundance changes at five and eleven hours

Protein levels develop from spores over germ bubbles to young mycelium in different ways. To categorize the different abundance behaviors we simply grouped the proteins into 4 clusters, based on the time point and their ratio to the starting point of the analysis. To define under which conditions a protein is up- or down-regulated we developed minimum limitation settings for the clusters. In Figure5-5A the theoretical settings are indicated. The protein abundance was transformed into the logarithm to the base 10 and the ration to the zero hour time point was defined for each time point. For each time point, five and eleven hour, the increased and decreased protein levels were added to the group and separated into solid and slight changes. Major changes were assumed when the protein abundance was altered by a factor of 10, which has a value of 1 in logarithmic scale. Protein abundances with minor changes were defined with an increase or decrease by the factor of 0.48 or higher, which is about all proteins that have an abundance alteration by a factor of three to ten. The average development of the different protein groups is shown in Figure5-5B. The alteration between the cluster A and B is not changing in a significant manner in relation to the average calculation. Both the up- and down-regulated proteins are mainly developing in the same way at five and eleven hours and on average all an abundance change higher then factor 10 at eleven hours. In Figure5-5C the total number of proteins that were grouped together is plotted. At five hours only about 300 proteins were up regulated and about 100 down-regulated. In contrast to later developmental stages of *A. gossypii* close to 800 proteins are increased in their abundance and more the 800 are decreased. In general the protein abundance changes are more dramatic over time and lead to the higher amount of protein variations at young mycelium compared to spores.

The top50 up-regulated proteins from spores to germ bubbles are listed in Table5-1 and their corresponding common names based on the *S. cerevisiae* data including the SGD description of first *S. cerevisiae* homolog (Table5-2). The total amounts of proteins that are induced by fold changes from three up to ca. 10,000 times are presented in the diagram in Figure5-6. All proteins that are induced more than three times (Figure5-6A) and the protein that are ten times and higher expressed at five hours (Figure5-6B) were analyzed with the DAVID Functional Annotation Bioinformatics tool (Huang da et al., 2009). The resulting categories display the score and general function. By comparing the frequencies of proteins, which share similar function in the cluster, with the frequency of randomly detection in the background of the total proteome, we calculated the probability (P-values) for the different functional annotations. The P-values are presenting the enrichment factor, the smaller the value the more this protein group is enriched in the cluster. The lists of the corresponding functional groups are presented below the diagrams in Figure5-6A and B. At five hours we find 321 proteins that are up-regulated combined in the cluster 1a. The functional annotations of the cluster show enrichment by the low P-values for example for ribosome biogenesis, nucleolus proteins and also 20 proteins that are involved in rRNA processing. Furthermore, regulation of transcription increased in proteins that are up regulated after five hours. Also the cell cycle proteins are significantly enriched with a P-value of 0.0036, because the P-Value is equal or smaller than 0.05, which is associated with a strong enrichment in this annotated class. The cluster 1b with a specification for higher induced proteins in germ bubbles show 166 proteins that are considered to be part of this category. Since in this category of cluster 1b fewer proteins were annotated, also the P-values are in general higher. With 19 proteins that are involved in the regulation of transcription the majority of the proteins is highly up-regulated. In contrast, cell cycle proteins are not significantly enriched in the cluster. Interestingly, all the proteins that are associated with the kinetochore are specifically located in the cluster of proteins that are highly induced during germinating after five hours.

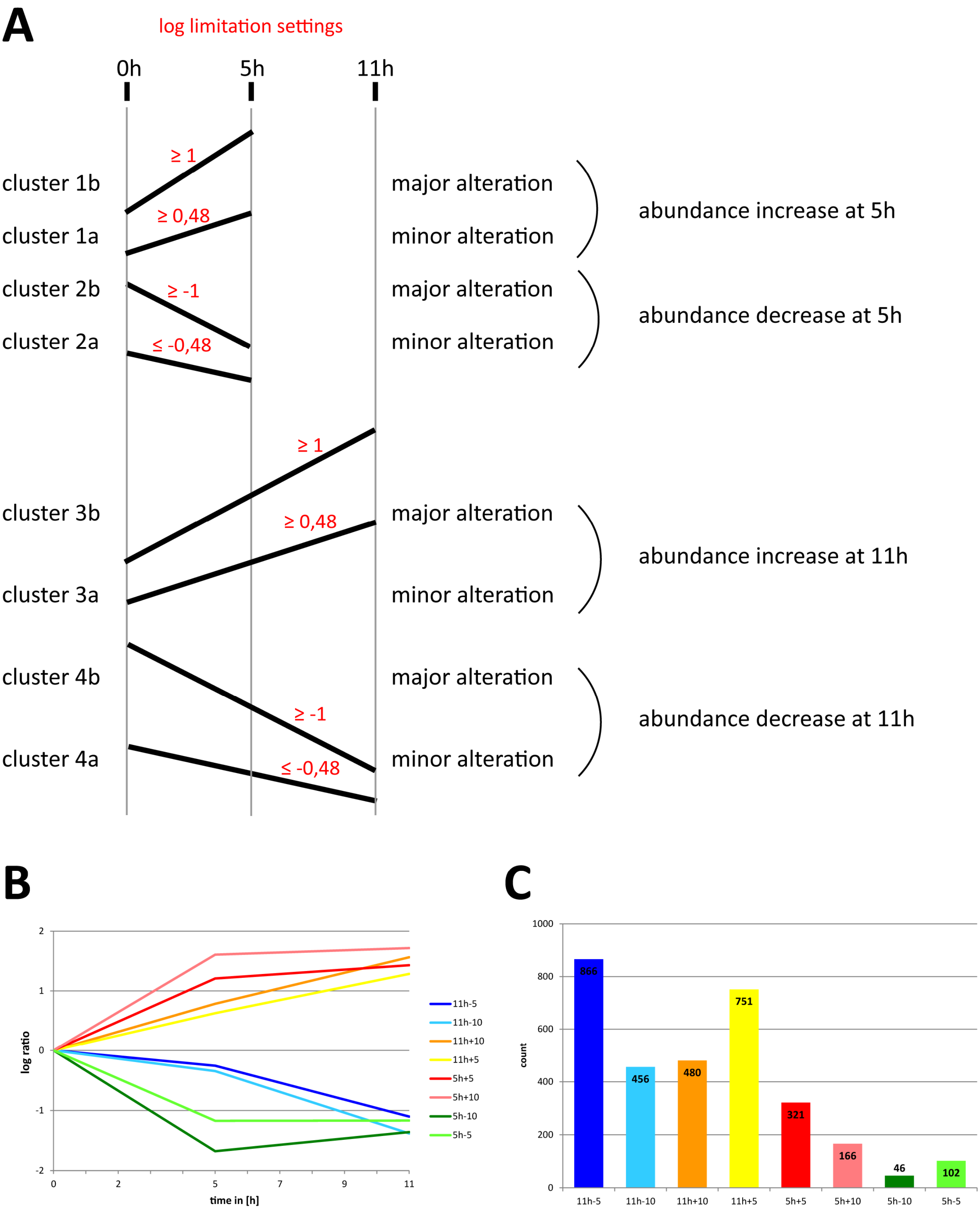


Figure5-5: Time point based cluster analyses. **(A)** Theoretical settings to group the proteins in four clusters. **(B)** Line diagram of the average log abundance of proteins from zero to eleven hours. **(C)** Column chart of protein numbers identify per cluster.

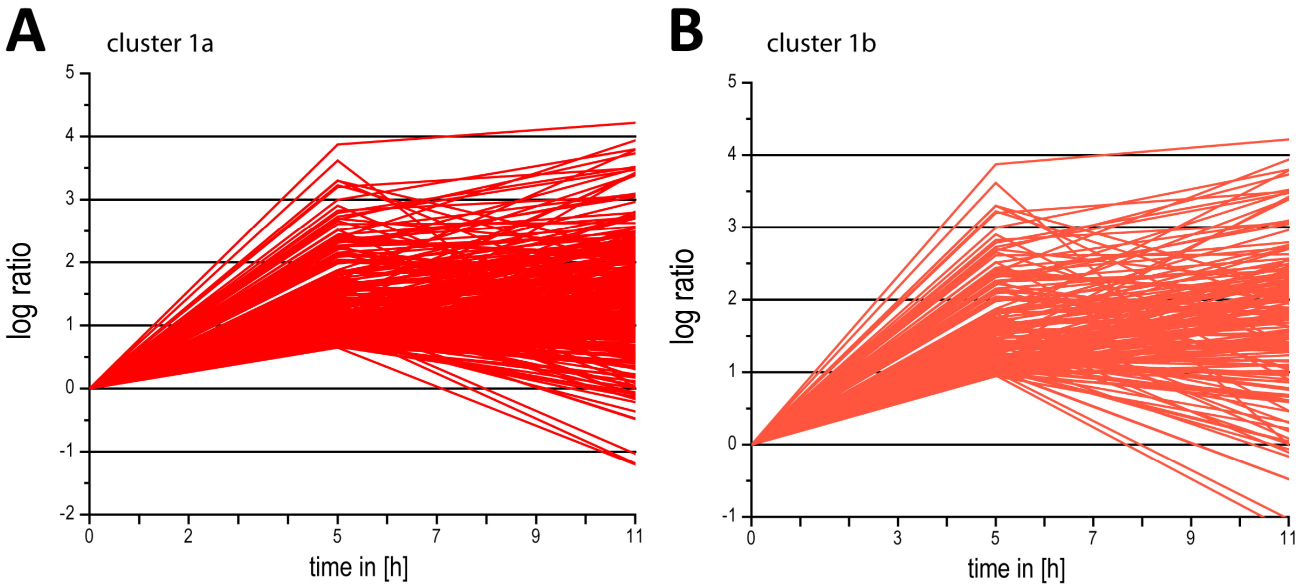
Table5-1: Ranking of the 50 most up-regulated proteins from zero to fife hours.

no.	accession	common name	0h relative abundance (all *1)	5h relative abundance (all *1)	11h relative abundance (all *1)	5h log ratio	11h log ratio	detected peptieds	possible peptides	protein size [aa]	protein size [MW]	Score
1	AER436C	Zrg8	0.0001	0.89	1.96	3.875	4.216	1	84	991	108977	45.77
2	AGR353C	YLR012C	0.0004	1.80	0.0004	3.618	0.000	1	23	177	21117	71.2
3	AFR212W	Asi1, Asi3	0.0003	0.61	0.007	3.302	1.381	1	33	626	70161	67.85
4	AEL294C	Ftr1	0.0040	7.86	0.59	3.298	2.175	3	15	366	40848	112.13
5	ACR193C	Nto1	0.0002	0.36	0.019	3.224	1.942	2	46	729	83068	106.41
6	AFR257W	Hap2	0.0003	0.49	0.98	3.199	3.497	1	32	268	29235	36.23
7	AFR679C	Spc25	0.0005	0.49	3.15	2.994	3.799	2	20	231	26060	55.13
8	ACR261C	YKR051W	0.0003	0.22	0.002	2.901	0.900	1	36	407	47446	58.24
9	AFR297W	Bas1	0.0010	0.65	1.11	2.822	3.054	1	42	743	82267	64.68
10	AGL168W	Cwc15	0.0004	0.27	1.44	2.786	3.520	1	23	184	19772	34
11	AGR049W	Ccw12	0.0025	1.33	7.46	2.725	3.475	1	4	132	12989	73.51
12	AFR014C	Hap3	0.0005	0.26	0.091	2.696	2.238	1	19	176	19356	35.1
13	AFR185C	Snf6	0.0008	0.37	4.17	2.677	3.735	2	13	284	32309	57.07
14	AEL257W	Gdt1	0.0006	0.25	0.014	2.653	1.405	1	21	277	30102	50.48
15	ABL021C	Vph2	0.0003	0.14	0.15	2.609	2.625	1	29	239	25740	40.12
16	ABR225C	Pet309	0.0002	0.077	0.001	2.517	0.468	1	104	943	108933	52.3
17	ADL308C	Mak3	0.0005	0.16	0.62	2.51	3.090	1	20	173	19822	49.15
18	AAR116W	Sme1	0.035	9.49	21.34	2.438	2.790	1	10	97	10985	63.66
19	ADR093W	NOHBY425	0.0006	0.15	0.056	2.421	2.004	1	18	194	20305	56.7
20	AAR089C	Esp1	0.0004	0.11	0.049	2.4	2.057	2	115	1618	186565	94.28
21	AFL127C	Ecm5	0.0001	0.027	0.97	2.382	3.938	1	89	1521	173608	60.48
22	AGR014W	Vps20	0.0003	0.081	0.040	2.373	2.063	1	29	219	24648	49.72
23	AFR044C	YHR132W-A, YN	0.062	14.15	32.14	2.36	2.717	2	6	126	13490	160.34
24	AER220C	Bud14	0.016	3.66	1.91	2.355	2.072	2	35	671	73804	111.11
25	AFR659W	Cwc23	0.0007	0.14	0.089	2.329	2.132	1	33	273	31808	45.34
26	ADR281W	Snu23	0.010	2.06	11.21	2.307	3.043	2	16	193	22139	144.11
27	AEL064C	Pet191	0.034	6.18	12.42	2.261	2.565	3	10	108	12376	199.1
28	AFR227W	YMR130W	0.043	7.78	0.47	2.253	1.039	1	27	271	29982	37.82
29	ACL197W	Prx1	0.0005	0.084	0.12	2.247	2.397	1	21	226	25626	42.66
30	ABL162C	Psy2	0.0014	0.24	1.77	2.235	3.096	4	47	838	94833	170.65
31	AFR673C	Slp1	0.0002	0.026	0.007	2.22	1.634	1	64	765	85683	35.11
32	AFR141C	Mus81	0.0005	0.080	1.56	2.201	3.491	2	69	604	68666	49.03
33	AER078W	YLR407W	0.0039	0.55	1.17	2.151	2.478	1	26	217	24255	64.83
34	AER166C	Ctf8	0.0007	0.094	0.038	2.149	1.756	1	15	133	14867	55.99
35	AGR258C	Tma7, Rbf7	0.0008	0.11	0.010	2.138	1.062	1	12	64	6853.7	48.49
36	ADR054C	Dem1	0.027	3.55	2.02	2.112	1.867	3	35	527	59019	200.52
37	AFR536W	NOHBY650	0.0003	0.037	0.10	2.085	2.489	1	115	1259	141261	36.16
38	AGR208W	Ump1	0.21	23.51	16.59	2.041	1.890	2	7	154	16734	82.92
39	ABR098C	Mcm1, Arg80	0.0061	0.66	0.39	2.029	1.799	2	5	206	22637	78.05
40	AGL262C	NOHBY713	0.039	3.92	2.54	2.004	1.816	4	21	325	36233	195.28
41	AGR207C	Swd3	0.0021	0.20	0.54	1.979	2.419	1	22	320	34767	57.91
42	AER038C	Spp381	0.0006	0.044	1.47	1.875	3.398	3	17	264	29163	103.96
43	AER341W	Mod5	0.027	1.96	6.33	1.856	2.365	2	26	442	51064	79.95
44	ABR004C	Aur1	0.0004	0.027	0.053	1.85	2.140	1	26	439	49420	38.33
45	AER127C	Ogg1	0.0002	0.014	0.12	1.841	2.754	1	48	391	44485	44.99
46	ADL034W	YGL176C	0.0084	0.56	0.58	1.829	1.841	1	42	449	47936	38.59
47	ADL398C	Cwp1	0.36	23.20	0.44	1.806	0.087	2	6	154	15877	71.58
48	AGR247C	Dsn1	0.0003	0.020	0.29	1.805	2.962	2	32	518	58589	85.47
49	AGR319W	Hip1	0.0092	0.59	1.99	1.805	2.337	4	33	626	68174	179.32
50	ACL105C	Sbe22, Sbe2	0.0009	0.054	0.30	1.799	2.538	2	46	814	89947	124.23

Table5-2: SGD based description of the 50 most up-regulated proteins from zero to five hours.

no.	common name	info
1	Zrg8	Protein of unknown function; authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies; GFP-fusion protein is localized to the cytoplasm; transcription induced under conditions of zinc deficiency
2	YLR012C	Putative protein of unknown function; YLR012C is not an essential gene
3	Asi1, Asi3	Putative integral membrane E3 ubiquitin ligase; acts with Asi2p and Asi3p to ensure the fidelity of SPS-sensor signalling by maintaining the dormant repressed state of gene expression in the absence of inducing signals
4	Ftr1	High affinity iron permease involved in the transport of iron across the plasma membrane; forms complex with Fet3p; expression is regulated by iron
5	Nto1	Subunit of the NuA3 histone acetyltransferase complex that acetylates histone H3; contains PHD finger domain that interacts with methylated histone H3
6	Hap2	Subunit of the heme-activated, glucose-repressed Hap2p/3p/4p/5p CCAAT-binding complex, a transcriptional activator and global regulator of respiratory gene expression; contains sequences sufficient for both complex assembly and DNA binding
7	Spc25	Component of the evolutionarily conserved kinetochore-associated Ndc80 complex (Ndc80p-Nuf2p-Spc24p-Spc25p); involved in chromosome segregation, spindle checkpoint activity and kinetochore clustering
8	YKR051W	Putative protein of unknown function
9	Bas1	Myb-related transcription factor involved in regulating basal and induced expression of genes of the purine and histidine biosynthesis pathways; also involved in regulation of meiotic recombination at specific genes
10	Cwc15	Non-essential protein involved in pre-mRNA splicing, component of a complex containing Cef1p; has similarity to S. pombe Cwf15p
11	Ccw12	Cell wall mannoprotein, mutants are defective in mating and agglutination, expression is downregulated by alpha-factor
12	Hap3	Subunit of the heme-activated, glucose-repressed Hap2p/3p/4p/5p CCAAT-binding complex, a transcriptional activator and global regulator of respiratory gene expression; contains sequences contributing to both complex assembly and DNA binding
13	Snf6	Subunit of the SWI/SNF chromatin remodeling complex involved in transcriptional regulation; functions interdependently in transcriptional activation with Snf2p and Snf5p
14	Gdt1	Putative protein of unknown function; expression is reduced in a gcr1 null mutant; GFP-fusion protein localizes to the vacuole; expression pattern and physical interactions suggest a possible role in ribosome biogenesis
15	Vph2	Integral membrane protein required for vacuolar H+-ATPase (V-ATPase) function, although not an actual component of the V-ATPase complex; functions in the assembly of the V-ATPase; localized to the endoplasmic reticulum (ER)
16	Pet309	Specific translational activator for the COX1 mRNA, also influences stability of intron-containing COX1 primary transcripts; localizes to the mitochondrial inner membrane; contains seven pentatricopeptide repeats (PPRs)
17	Mak3	Catalytic subunit of N-terminal acetyltransferase of the NatC type; required for replication of dsRNA virus
18	Sme1	Core Sm protein Sm E; part of heteroheptameric complex (with Smb1p, Smd1p, Smd2p, Smd3p, Smx3p, and Smx2p) that is part of the spliceosomal U1, U2, U4, and U5 snRNPs; homolog of human Sm E
19	NOHBY425	
20	Esp1	Separase, a caspase-like cysteine protease that promotes sister chromatid separation by mediating dissociation of the cohesin Scc1p from chromatin; inhibits protein phosphatase 2A-Cdc55p to promote mitotic exit; inhibited by Pds1p
21	Ecm5	Non-essential protein of unknown function, contains ATP/GTP-binding site motif A; null mutant exhibits cellular volume up to four times greater than wild-type, also large drooping buds with elongated necks
22	Vps20	Myristoylated subunit of ESCRTIII, the endosomal sorting complex required for transport of transmembrane proteins into the multivesicular body pathway to the lysosomal/vacuolar lumen; cytoplasmic protein recruited to endosomal membranes
23	YHR132W-A, YNL157W	Protein required for initiation of G0 program; prevents degradation of nutrient-regulated mRNAs via the 5'-3' mRNA decay pathway; phosphorylated by Rim15p; GFP protein localizes to the cytoplasm and nucleus; similar to Igo1p
24	Bud14	Protein involved in bud-site selection, Bud14p-Glc7p complex is a cortical regulator of dynein; inhibitor of the actin assembly factor Bnr1p (formin); diploid mutants display a random budding pattern instead of the wild-type bipolar pattern
25	Cwc23	Component of a complex containing Cef1p, putatively involved in pre-mRNA splicing; has similarity to E. coli DnaJ and other DnaJ-like proteins and to S. pombe Cwf23p
26	Snu23	Component of U4/U6.U5 snRNP involved in mRNA splicing via spliceosome
27	Pet191	Protein required for assembly of cytochrome c oxidase; exists as an oligomer that is integral to the mitochondrial inner membrane and faces the intermembrane space; contains a twin Cx9C motif
28	YMR130W	Putative protein of unknown function; YMR130W is not an essential gene
29	Prx1	Mitochondrial peroxiredoxin (1-Cys Prx) with thioredoxin peroxidase activity, has a role in reduction of hydroperoxides; reactivation requires Trr2p and glutathione; induced during respiratory growth and oxidative stress; phosphorylated
30	Psy2	Putative subunit of an evolutionarily conserved protein phosphatase complex containing the catalytic subunit Pph3p and the regulatory subunit Psy4p; required for cisplatin and oxaliplatin resistance; putative homolog of mammalian R3
31	Slp1	Integral membrane protein of unknown function; member of the SUN-like family of proteins; genetic interactions suggest a role in folding of ER membrane proteins
32	Mus81	Subunit of the structure-specific Mms4p-Mus81p endonuclease that cleaves branched DNA; involved in DNA repair, replication fork stability, and joint molecule formation/resolution during meiotic recombination; helix-hairpin-helix protein
33	YLR407W	Putative protein of unknown function; null mutant displays elongated buds and a large fraction of budded cells have only one nucleus
34	Ctf8	Subunit of a complex with Ctf18p that shares some subunits with Replication Factor C and is required for sister chromatid cohesion
35	Tma7, Rbf7	Protein of unknown that associates with ribosomes; null mutant exhibits translation defects, altered polyribosome profiles, and resistance to the translation inhibitor anisomcyin
36	Dem1	Mitochondrial 5'-3' exonuclease and sliding exonuclease, required for mitochondrial genome maintenance; distantly related to the RecB nuclease domain of bacterial RecBCD recombinases; may be regulated by the transcription factor Ace2
37	NOHBY650	
38	Ump1	Short-lived chaperone required for correct maturation of the 20S proteasome; may inhibit premature dimerization of proteasome half-mers; degraded by proteasome upon completion of its assembly
39	Mcm1, Arg80	Transcription factor involved in cell-type-specific transcription and pheromone response; plays a central role in the formation of both repressor and activator complexes
40	NOHBY713	
41	Swd3	Essential subunit of the COMPASS (Set1C) complex, which methylates histone H3 on lysine 4 and is required in transcriptional silencing near telomeres; WD40 beta propeller superfamily member and ortholog of mammalian WDR5
42	Spp381	mRNA splicing factor, component of U4/U6.U5 tri-snRNP; interacts genetically and physically with Prp38p
43	Mod5	Delta 2-isopentenyl pyrophosphate:tRNA isopentenyl transferase, required for biosynthesis of the modified base isopentenyladenosine in mitochondrial and cytoplasmic tRNAs; gene is nuclear and encodes two isozymic forms
44	Aur1	Phosphatidylinositol:ceramide phosphoinositol transferase (IPC synthase), required for sphingolipid synthesis; can mutate to confer aureobasidin A resistance
45	Ogg1	Mitochondrial glycosylase/lyase that specifically excises 7,8-dihydro-8-oxoguanine residues located opposite cytosine or thymine residues in DNA, repairs oxidative damage to mitochondrial DNA, contributes to UVA resistance
46	YGL176C	Putative protein of unknown function; deletion mutant is viable and has no detectable phenotype
47	Cwp1	Cell wall mannoprotein that localizes specifically to birth scars of daughter cells, linked to a beta-1,3- and beta-1,6-glucan heteropolymer through a phosphodiester bond; required for propionic acid resistance
48	Dsn1	Essential component of the MIND kinetochore complex (Mtw1p Including Nnf1p-Nsl1p-Dsn1p) which joins kinetochore subunits contacting DNA to those contacting microtubules; important for chromosome segregation
49	Hip1	High-affinity histidine permease, also involved in the transport of manganese ions
50	Sbe22, Sbe2	Protein involved in the transport of cell wall components from the Golgi to the cell surface; similar in structure and functionally redundant with Sbe2p; involved in bud growth

Only 103 proteins are decreased in their abundance levels after five hours (Figure5-5C). The top 50 highest down-regulated proteins are presented in Table5-3 and the corresponding function is listed in Table5-4. The cluster 2a and b are represented with a low number of proteins (Figure5-7). In the DAVID Functional Annotation only a few clusters were computed. Significantly enriched are glycoproteins and integral to membrane proteins, which are proteins that are embedded in the bilayer with the whole protein or part of the peptide. A putative mating-type transcription factor is also significantly reduced after five hours of growth in cluster 2a. With a higher enrichment factor based on the lower P-value the functional group in cluster 2b is more significant, which characterizes proteins that are decreased in abundance by more than the 10 times. No other significant functional annotations were discovered in the cluster 2b that contains only 47 proteins (Figure5-7B).



function - cluster 1a	count	enrichment score	P-Value
nucleolus	20	7.16	0.00000000044
ribosome biogenesis	24	7.16	0.0000000011
rRNA processing	20	7.16	0.000000013
Pyrimidine metabolism	13	1.5	0.000019
DNA replication	7	1.5	0.0023
chromosome	14	1.43	0.0019
cell cycle phase (M phase)	10	1.43	0.0036
regulation of transcription	27	1.45	0.0054
Purine metabolism	10	1.5	0.0086
kinetochore	5	1.43	0.019
total	**321**		

function - cluster 1b	count	enrichment score	P-Value
regulation of transcription	19	2.09	0.00026
kinetochore	5	1.65	0.0015
RNA processing	12	1.53	0.033
Spliceosome	6	1.53	0.0021
Pyrimidine metabolism	6	0.98	0.0078
Purine metabolism	4	0.98	0.21
total	**166**		

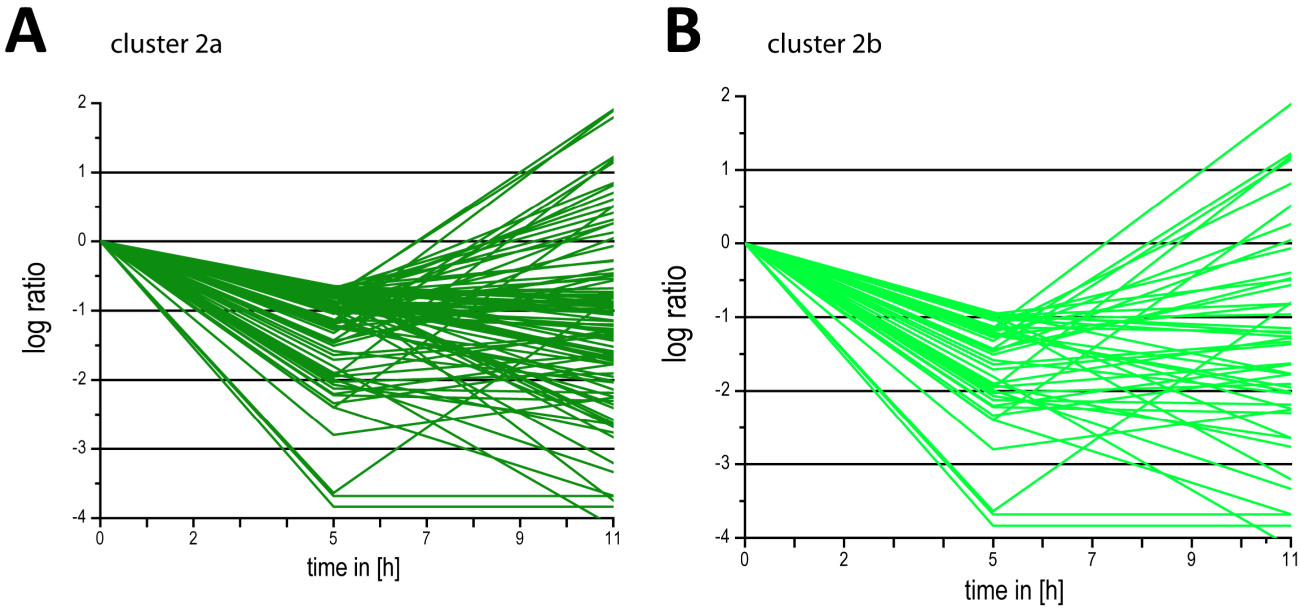
Figure5-6: All up-regulated proteins in five hour germ bubbles. **(A)** Minor increase of protein abundances. List of biological groups of this cluster is shown below. **(B)** Proteins that show a major increase in their abundances. Functional annotation of the increased protein groups is listed below.

Table5-3: Ranking of the 50 most down regulated proteins from zero to fife hours.

no.	accession	common name	0h relative abundance (all *1)	5h relative abundance (all *1)	11h relative abundance (all *1)	5h log ratio	11h log ratio	detected peptieds	possible peptides	protein size [aa]	protein size [MW]	Score
1	ACL125C	Lsm8	17.15	0.0025	0.0025	-3.836	-3.8363	1	4	106	11421	56.17
2	AFR522C	Ste2	1.78	0.0004	0.0004	-3.681	-3.681	1	27	458	50476	67.68
3	AER029C	Atg19	0.92	0.0002	0.14	-3.638	-0.8062	2	47	565	60964	60.59
4	AGR408W		0.78	0.0013	0.0044	-2.798	-2.2524	1	8	172	19278	78.44
5	AFL200W	Cep3	0.040	0.0002	0.13	-2.399	0.51043	1	62	596	68869	55.04
6	AGR225W		11.93	0.048	0.0025	-2.395	-3.6787	1	4	94	10083	50.37
7	ADR300C	Ksp1	0.028	0.0001	0.0005	-2.334	-1.7739	1	76	890	100702	53.63
8	AGL365C	NOHBY725	0.11	0.0006	0.0005	-2.221	-2.3012	1	46	599	67207	43.61
9	AAL151C	Emc6	17.46	0.11	0.40	-2.201	-1.6411	2	5	106	11832	118.49
10	AAL184W	YML108W	1.71	0.013	0.011	-2.123	-2.186	1	9	112	12810	35.11
11	AGL002C	Snf8	0.29	0.0025	0.0007	-2.069	-2.6442	1	15	232	26842	39.18
12	ACR156W	Rny1	2.17	0.020	0.027	-2.027	-1.9057	3	20	292	33394	91.95
13	AAL004W	Ldb16	0.80	0.0082	0.0004	-1.990	-3.3339	2	27	311	35020	55.48
14	AFR736C	Kar4	0.027	0.0003	0.41	-1.955	1.18769	1	34	322	37140	36.16
15	ADL393W	NOHBY421	0.041	0.0005	0.0010	-1.948	-1.6273	1	22	142	16602	34.86
16	AMI008W	Cob	5.19	0.063	0.0089	-1.916	-2.7639	1	7	384	43802	36.48
17	ADR344W	Nse1	0.031	0.0004	0.0016	-1.896	-1.2821	1	25	281	31082	47.03
18	AGR359C	Ppe1	5.26	0.082	0.0004	-1.805	-4.1187	1	25	385	42796	47.31
19	ADR051C	Dur1,2	0.020	0.0004	0.0008	-1.708	-1.3667	1	163	1819	197447	36.47
20	ADL333C	Uba3	0.076	0.0017	0.0036	-1.640	-1.3306	1	23	319	34057	39.68
21	AAL089W	YOR246C	23.17	0.61	0.21	-1.583	-2.0357	3	14	328	37370	136.04
22	AGL155W	Taf10	26.65	0.83	9.00	-1.507	-0.4713	3	9	198	22316	168.67
23	AFL136W	YPL230W, Rgm1	0.40	0.014	0.44	-1.467	0.04298	1	36	391	41090	60.59
24	AGR295C	Atp14	78.74	2.87	21.47	-1.438	-0.5643	2	3	117	12312	44.93
25	AFR728W	Kre28	0.37	0.014	5.11	-1.428	1.14336	4	27	361	41185	188.44
26	AAR029W	Dam1	0.026	0.0013	0.44	-1.322	1.22203	1	35	339	36658	43.85
27	AFR180C	Rrn10	5.48	0.29	9.99	-1.275	0.26077	1	11	134	14574	48.42
28	AEL297W	Rad54	1.55	0.088	0.23	-1.248	-0.826	2	61	895	100258	67.69
29	AER096C	Cnb1	115.53	6.68	1.29	-1.238	-1.9507	4	12	175	19713	212.81
30	AER287W	Mdm1	0.038	0.0023	0.0004	-1.227	-2.013	2	82	1153	133676	82.96
31	ADR238C	NOHBY432	53.68	3.29	0.90	-1.213	-1.7753	4	22	476	53661	164.65
32	AGR349W	Mcm21	0.48	0.031	0.19	-1.189	-0.3979	2	22	361	41336	130.8
33	ABL097C	Mic14	1.46	0.10	9.50	-1.165	0.81471	3	10	119	13750	105.76
34	ADR068W	Clb4, Clb3	0.0036	0.0003	0.28	-1.146	1.8939	1	39	375	41446	51.14
35	AER256C	Kei1	8.16	0.59	0.52	-1.138	-1.1983	2	12	239	27567	123.51
36	AGR326C	YJR011C	4.94	0.37	0.028	-1.126	-2.2412	1	10	202	23132	37.33
37	AER431C	Vhs2, Mlf3	3.50	0.32	2.98	-1.043	-0.0696	3	22	418	43920	154.1
38	AEL185C	Chk1	2.82	0.26	0.31	-1.028	-0.964	2	37	515	57863	77.62
39	AFR729C	Cab1	5.16	0.49	0.28	-1.021	-1.2624	3	23	352	38424	133.39
40	AER125C	Lsm3	10.78	1.03	0.0067	-1.019	-3.2058	1	10	83	9167	54.29
41	AFR604C	Spo21	0.13	0.014	0.0003	-0.994	-2.6433	2	56	844	96461	103.04
42	AFL188C	Dun1	10.33	1.08	0.73	-0.981	-1.1519	3	29	472	53766	80.38
43	ADL368W	NOHBY420	2.16	0.23	0.33	-0.980	-0.8195	3	23	314	35320	92.67
44	ACR007W	Tsr3	36.19	3.99	11.49	-0.957	-0.4981	4	16	313	35484	128.12
45	AAR040C	Ccz1	2.37	0.26	0.14	-0.952	-1.2193	4	44	663	76266	179.55
46	ABL171W	Bre4	1.23	0.14	0.021	-0.935	-1.7672	2	55	1053	118951	83.99
47	AFR092W	Bck1	0.24	0.028	0.42	-0.926	0.25701	4	79	1423	158043	124.46
48	AFL223W	Cyb5	1086.35	128.99	28.26	-0.925	-1.5848	4	7	165	18221	283.56
49	AGL011C	NOHBY701	5.95	0.72	0.026	-0.916	-2.3518	1	17	317	34779	63.52
50	AFR177C	NOHBY630	0.090	0.012	0.45	-0.891	0.70377	2	38	363	40895	49.06

Table5-4: SGD based description of the 50 highest down regulated proteins from zero to five hours.

no.	common name	info
1	Lsm8	Lsm (Like Sm) protein; forms heteroheptameric complex (with Lsm2p, Lsm3p, Lsm4p, Lsm5p, Lsm6p, and Lsm7p) that is part of spliceosomal U6 snRNP and is also implicated in processing of pre-tRNA, pre-snoRNA, and pre-rRNA
2	Ste2	Receptor for alpha-factor pheromone; seven transmembrane-domain GPCR that interacts with both pheromone and a heterotrimeric G protein to initiate the signaling response that leads to mating between haploid a and alpha cells
3	Atg19	Receptor protein specific for the cytoplasm-to-vacuole targeting (Cvt) pathway; delivers cargo proteins aminopeptidase I (Lap4p) and alpha-mannosidase (Ams1p) to the phagophore assembly site for packaging into Cvt vesicles
4	AGR408W	
5	Cep3	Essential kinetochore protein, component of the CBF3 complex that binds the CDEIII region of the centromere; contains an N-terminal Zn2Cys6 type zinc finger domain, a C-terminal acidic domain, and a putative coiled coil dimerization domain
6	AGR225W	
7	Ksp1	Ser/thr protein kinase; nuclear translocation required for haploid filamentous growth; regulates filamentous growth induced nuclear translocation of Bcy1p, Fus3p, and Sks1p; overproduction causes allele-specific suppression of prp20-10
8	NOHBY725	
9	Emc6	Member of a transmembrane complex required for efficient folding of proteins in the ER; null mutant displays induction of the unfolded protein response
10	YML108W	Putative protein of unknown function whose structure defines a new subfamily of the split beta-alpha-beta sandwiches; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm and nucleus; YML108W is not an essential gene
11	Snf8	Component of the ESCRT-II complex, which is involved in ubiquitin-dependent sorting of proteins into the endosome; appears to be functionally related to SNF7; involved in glucose derepression
12	Rny1	Vacuolar RNase of the T(2) family, relocalizes to the cytosol where it cleaves tRNAs upon oxidative or stationary phase stress; promotes apoptosis under stress conditions and this function is independent of its catalytic activity
13	Ldb16	Protein of unknown function; null mutants have decreased net negative cell surface charge; GFP-fusion protein expression is induced in response to the DNA-damaging agent MMS; native protein is detected in purified mitochondria
14	Kar4	Transcription factor required for gene regulation in response to pheromones; also required during meiosis; exists in two forms, a slower-migrating form more abundant during vegetative growth and a faster-migrating form induced by pheromone
15	NOHBY421	
16	Cob	
17	Nse1	Essential subunit of the Mms21-Smc5-Smc6 complex; nuclear protein required for DNA repair and growth; has a nonstructural role in the maintenance of chromosomes
18	Ppe1	Protein with carboxyl methyl esterase activity that may have a role in demethylation of the phosphoprotein phosphatase catalytic subunit; also identified as a small subunit mitochondrial ribosomal protein
19	Dur1,2	Urea amidolyase, contains both urea carboxylase and allophanate hydrolase activities, degrades urea to CO2 and NH3; expression sensitive to nitrogen catabolite repression and induced by allophanate, an intermediate in allantoin degradation
20	Uba3	Protein that acts together with Ula1p to activate Rub1p before its conjugation to proteins (neddylation), which may play a role in protein degradation; GFP-fusion protein localizes to the cytoplasm in a punctate pattern
21	YOR246C	Protein with similarity to oxidoreductases, found in lipid particles; required for replication of Brome mosaic virus in S. cerevisiae, a model system for studying replication of positive-strand RNA viruses in their natural hosts
22	Taf10	Subunit (145 kDa) of TFIID and SAGA complexes, involved in RNA polymerase II transcription initiation and in chromatin modification
23	YPL230W, Rgm1	Putative transcription factor containing a C2H2 zinc finger; mutation affects transcriptional regulation of genes involved in growth on non-fermentable carbon sources, response to salt stress and cell wall biosynthesis
24	Atp14	Subunit h of the F0 sector of mitochondrial F1F0 ATP synthase, which is a large, evolutionarily conserved enzyme complex required for ATP synthesis
25	Kre28	Subunit of a kinetochore-microtubule binding complex with Spc105p that bridges centromeric heterochromatin and kinetochore MAPs and motors, and is also required for sister chromatid bi-orientation and kinetochore binding of SAC components
26	Dam1	Essential subunit of the Dam1 complex (aka DASH complex), couples kinetochores to the force produced by MT depolymerization thereby aiding in chromosome segregation; Ipl1p target for regulating kinetochore-MT attachments
27	Rrn10	Protein involved in promoting high level transcription of rDNA, subunit of UAF (upstream activation factor) for RNA polymerase I
28	Rad54	DNA-dependent ATPase, stimulates strand exchange by modifying the topology of double-stranded DNA; involved in the recombinational repair of double-strand breaks in DNA during vegetative growth and meiosis; member of the SWI/SNF family
29	Cnb1	Calcineurin B; the regulatory subunit of calcineurin, a Ca++/calmodulin-regulated type 2B protein phosphatase which regulates Crz1p (a stress-response transcription factor), the other calcineurin subunit is encoded by CNA1 and/or CMP1
30	Mdm1	Intermediate filament protein, required for nuclear and mitochondrial transmission to daughter buds; contains a Phox homology (PX) domain and specifically binds phosphatidylinositol 3-phosphate (PtdIns-3-P)
31	NOHBY432	
32	Mcm21	Protein involved in minichromosome maintenance; component of the COMA complex (Ctf19p, Okp1p, Mcm21p, Ame1p) that bridges kinetochore subunits that are in contact with centromeric DNA and the subunits bound to microtubules
33	Mic14	Mitochondrial intermembrane space protein, required for normal oxygen consumption; contains twin cysteine-x9-cysteine motifs
34	Clb4, Clb3	B-type cyclin involved in cell cycle progression; activates Cdc28p to promote the G2/M transition; may be involved in DNA replication and spindle assembly; accumulates during S phase and G2, then targeted for ubiquitin-mediated degradation
35	Kei1	Component of inositol phosphorylceramide (IPC) synthase; forms a complex with Aur1p and regulates its activity; required for IPC synthase complex localization to the Golgi; post-translationally processed by Kex2p; KEI1 is an essential gene
36	YJR011C	Putative protein of unknown function; GFP-fusion protein expression is induced in response to the DNA-damaging agent MMS
37	Vhs2, Mlf3	Cytoplasmic protein of unknown function; identified as a high-copy suppressor of the synthetic lethality of a sis2 sit4 double mutant, suggesting a role in G1/S phase progression; similar to Mlf3p
38	Chk1	Serine/threonine kinase and DNA damage checkpoint effector, mediates cell cycle arrest via phosphorylation of Pds1p; phosphorylated by checkpoint signal transducer Mec1p; homolog of S. pombe and mammalian Chk1 checkpoint kinase
39	Cab1	Pantothenate kinase (ATP:D-pantothenate 4'-phosphotransferase, EC 2.7.1.33); catalyzes the first committed step in the universal biosynthetic pathway for synthesis of coenzyme A (CoA)
40	Lsm3	Lsm (Like Sm) protein; part of heteroheptameric complexes (Lsm2p-7p and either Lsm1p or 8p): cytoplasmic Lsm1p complex involved in mRNA decay; nuclear Lsm8p complex part of U6 snRNP and possibly involved in processing tRNA, snoRNA, and rRNA
41	Spo21	Component of the meiotic outer plaque of the spindle pole body, involved in modifying the meiotic outer plaque that is required prior to prospore membrane formation
42	Dun1	Cell-cycle checkpoint serine-threonine kinase required for DNA damage-induced transcription of certain target genes, phosphorylation of Rad55p and Sml1p, and transient G2/M arrest after DNA damage; also regulates postreplicative DNA repair
43	NOHBY420	
44	Tsr3	Protein required for correct processing of the 20S pre-rRNA at site D to generate mature 18S rRNA; green fluorescent protein (GFP)-fusion protein localizes to both the cytoplasm and the nucleus
45	Ccz1	Protein involved in vacuolar assembly, essential for autophagy and the cytoplasm-to-vacuole pathway
46	Bre4	Zinc finger protein containing five transmembrane domains; null mutant exhibits strongly fragmented vacuoles and sensitivity to brefeldin A, a drug which is known to affect intracellular transport
47	Bck1	Mitogen-activated protein (MAP) kinase kinase kinase acting in the protein kinase C signaling pathway, which controls cell integrity; upon activation by Pkc1p phosphorylates downstream kinases Mkk1p and Mkk2p
48	Cyb5	Cytochrome b5, involved in the sterol and lipid biosynthesis pathways; acts as an electron donor to support sterol C5-6 desaturation
49	NOHBY701	
50	NOHBY630	

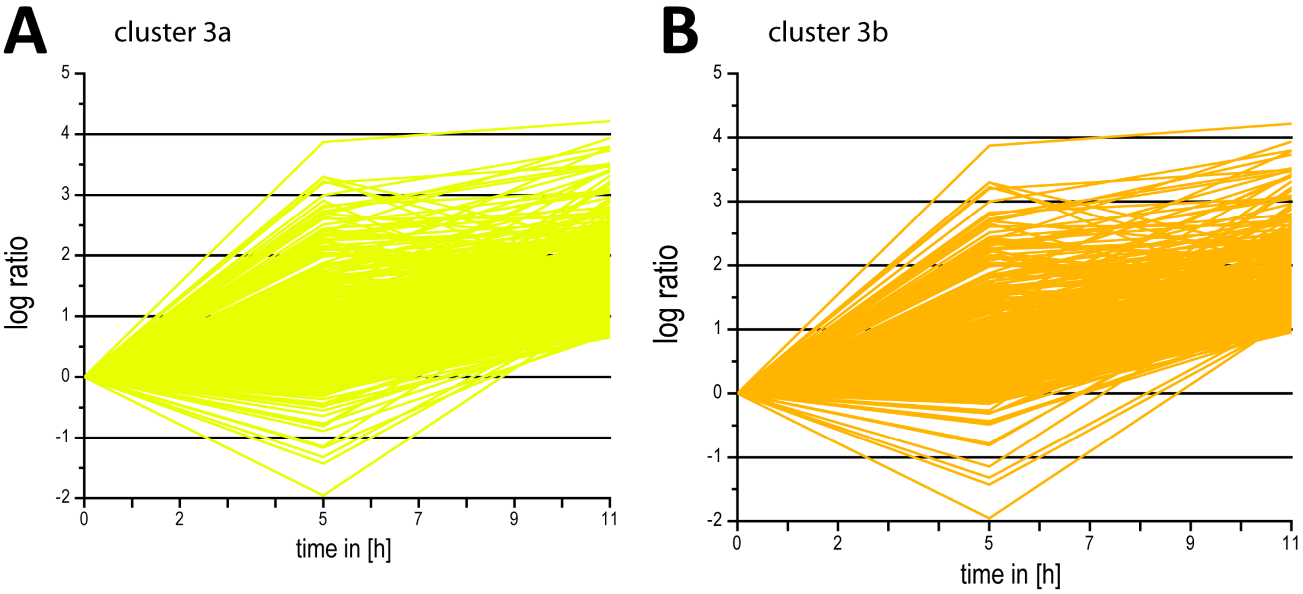


function	count	enrichment score	P-Value
EF-HAND 2	4	1.61	0.0033
glycoprotein	5	1.45	0.0088
integral to membrane	12	1.45	0.017
ATP binding	13	0.82	0.088
Putative mating-type transcription factor	3	0.64	0.0015
metal ion binding	13	0.52	0.191
total	**102**		

function	count	enrichment score	P-Value
Putative mating-type transcription	3	1.03	0.00032
ATP binding	7	0.8	0.067
total	**47**		

Figure5-7: All down regulated proteins in five hour germ bubbles. **(A)** Minor decrease in the protein abundance with the list of biological decreased protein groups in this cluster. **(B)** Only proteins that are reduced by a major value including the functional annotation of decreased protein groups.

A young mycelium of *A. gossypii* is morphologically much more altered compared to five hours stage where still a high amount of spores is present in the cell mix. After eleven hours the changes in protein abundances and composition are more dynamic then after five hours. Table5-5 shows the top 50 up-regulated proteins after eleven hours of growth under standard conditions. The cellular functions of the top 50 proteins are summarized in Table5-6. The clusters of the induced proteins during germination are plotted in Figure5-8. The characterization of the protein development is presented in the diagram. The logarithms to the base 10 values are placed in the relation to the time in hours and separated into tow clusters. Cluster 3a up-regulated proteins with an abundance increase of more than log 5 (Figure5-8A) and cluster 3b that contains only proteins that are induced with a higher fold change of log 1 (Figure5-8B). In total we sorted 751 proteins in the first group and by computing with the DAVID Functional Annotation Bioinformatics tool we were able to identify several biological sets of proteins. The highest enriched cellular components are the intracellular non-membrane-bounded organelles with a total number of 121 proteins and an P-value of $3.7x10^{-25}$. The proteins of ribonucleoprotein biogenesis are significantly enriched with a P-value of $1.1x10^{-28}$ and 59 proteins could be defined in this cluster. Proteins that are imported for DNA replication are also significantly increased in this cluster. This correlates well with the enrichment of proteins that are involved in the pyrimidine and purine metabolism. Both have a clearly low P-value which indicates the high association of this protein groups with growth. The rapid nuclear duplication during the germination of *A. gossypii* demands a high increase in the amount of proteins involved in this process in order to produce the high amount of biomass. Similarly, the RNA polymerase shows a high significance to be induced during this development, which is important to proportionally increase the transcription of genes with the increase of DNA. Like already shown at five hours the kinetochore proteins are also in a significant manner induced after eleven hours. We were able to identify in this cluster with a high coverage, 14 proteins of the total 19 kinetochore associated proteins. In the cluster 3b we could detect the same functional groups like in cluster 3a. With overall 481 proteins that are induced by a factor of 10 or more, we found a high number of proteins for these functional categories. This indicates the dramatic increase in the expression of the protein groups and their central role in the biomass increase and the growth of *A. gossypii*.



function - cluster 3a	count	enrichment score	P-Value
ribonucleoprotein complex biogenesis	59	19.13	1.10E-28
intracellular non-membrane-bounded org	121	10.13	3.70E-25
Pyrimidine metabolism	26	4.5	0.000000026
RNA polymerase	16	4.5	0.00000030
Purine metabolism	23	4.5	0.000096
regulation of transcription	65	3.02	0.0000059
ribonucleoprotein complex	14	1.71	0.015
cell division	26	1.71	0.90
kinetochore	14	1.26	0.0022
DNA replication	17	1.26	0.0071
total	**751**		

function - cluster 3b	count	enrichme	P-Value
ribonucleoprotein complex biogenesis	44	15.1	1.20E-22
intracellular non-membrane-bounded organelle	76	15.1	1.00E-15
Pyrimidine metabolism	18	3.6	0.00000065
RNA polymerase	11	3.6	0.000015
Purine metabolism	16	3.6	0.00021
regulation of transcription	44	3.31	0.000045
ribonucleoprotein complex	43	2.49	0.0000037
cell division	12	1.81	0.0005
kinetochore	7	1.81	0.0032
DNA replication	7	1.01	0.018
total	**481**		

Figure5-8: All up regulated proteins in eleven hour young mycelium. **(A)** Minor increase in the protein abundance with the list of biological increased protein groups in this cluster. **(B)** Only proteins that are induced by a major value including the functional annotation of increased protein groups.

Table5-5: Ranking of the 50 most up regulated proteins from zero to eleven hours.

no.	accession	common name	0h relative abundance (all *1)	5h relative abundance (all *1)	11h relative abundance (all *1)	5h log ratio	11h log ratio	detected peptieds	possible peptides	protein size [aa]	protein size [MW]	Score
1	AER436C	Zrg8	0.0001	0.89	1.96	3.875	4.216	1	84	991	108977	46
2	AFL127C	Ecm5	0.0001	0.027	0.97	2.382	3.938	1	89	1521	173608	60
3	AFR679C	Spc25	0.0005	0.49	3.15	2.994	3.799	2	20	231	26060	55
4	AFR743C		0.0029	0.15	17.93	1.718	3.787	1	5	101	10828	70
5	AFR185C	Snf6	0.0008	0.37	4.17	2.677	3.735	2	13	284	32309	57
6	AGL168W	Cwc15	0.0004	0.27	1.44	2.786	3.520	1	23	184	19772	34
7	AFR257W	Hap2	0.0003	0.49	0.98	3.199	3.497	1	32	268	29235	36
8	AFR141C	Mus81	0.0005	0.080	1.56	2.201	3.491	2	69	604	68666	49
9	AGR049W	Ccw12	0.0025	1.33	7.46	2.725	3.475	1	4	132	12989	74
10	ACR185W	Fmp26	0.0017	0.009	4.62	0.733	3.427	2	9	98	11528	97
11	ADL318C	Mms4	0.0001	0.002	0.35	1.272	3.423	1	76	714	79295	57
12	AER038C	Spp381	0.0006	0.044	1.47	1.875	3.398	3	17	264	29163	104
13	ABR083C	Cse4	0.0003	0.008	0.79	1.415	3.387	1	31	204	23887	82
14	ADL151W	Pex21, Pex18	0.0008	0.0010	1.56	0.109	3.315	2	17	307	32960	89
15	AFL140W	Nsl1	0.0004	0.0004	0.66	0.000	3.219	1	25	217	24711	42
16	AEL222C	Cin5, Yap6	0.0013	0.0013	1.87	0.000	3.175	2	8	234	24859	55
17	ABL162C	Psy2	0.0014	0.24	1.77	2.235	3.096	4	47	838	94833	171
18	ADL308C	Mak3	0.0005	0.16	0.62	2.510	3.090	1	20	173	19822	49
19	AFR297W	Bas1	0.0010	0.65	1.11	2.822	3.054	1	42	743	82267	65
20	ADR281W	Snu23	0.010	2.06	11.21	2.307	3.043	2	16	193	22139	144
21	AGR247C	Dsn1	0.0003	0.020	0.29	1.805	2.962	2	32	518	58589	85
22	AAR161W	Duo1	0.0017	0.0009	1.48	-0.268	2.943	1	11	194	21470	45
23	ADR133W	YKL075C	0.0004	0.0004	0.33	0.000	2.923	1	25	415	47802	48
24	AAR011C	YOL036W, YIR016W	0.0002	0.0002	0.13	0.000	2.884	1	57	747	79237	49
25	AFR224C	Smf3	0.0003	0.0003	0.23	0.000	2.873	1	32	472	52218	35
26	AEL319C	Slx9	0.0003	0.003	0.20	0.923	2.805	1	32	188	21494	38
27	AAR116W	Sme1	0.035	9.49	21.34	2.438	2.790	1	10	97	10985	64
28	AER127C	Ogg1	0.0002	0.014	0.12	1.841	2.754	1	48	391	44485	45
29	AFR118W	Ndl1	0.0005	0.0005	0.28	0.000	2.749	1	20	232	25558	49
30	AEL292W	Rec8	0.0078	0.36	4.10	1.664	2.720	3	45	629	69832	127
31	AFR044C	YHR132W-A, YNL157W	0.062	14.15	32.14	2.360	2.717	2	6	126	13490	160
32	ACR237C	Dot6	0.31	9.80	153.79	1.503	2.699	8	28	465	51450	426
33	ABR150C	Nnf1	0.0003	0.0003	0.15	0.000	2.661	1	30	199	22744	41
34	ABL021C	Vph2	0.0003	0.14	0.15	2.609	2.625	1	29	239	25740	40
35	AFR611C	Shg1	0.0005	0.0005	0.21	0.000	2.598	1	19	127	14094	58
36	AEL064C	Pet191	0.034	6.18	12.42	2.261	2.565	3	10	108	12376	199
37	AEL260C	Nut2	0.040	0.43	14.27	1.028	2.554	2	9	151	17341	72
38	AEL180C	Irc8	0.0015	0.014	0.53	0.983	2.551	1	57	650	69402	76
39	AEL216C	Mga1	0.22	0.47	78.49	0.322	2.548	5	18	450	48897	197
40	ACL105C	Sbe22, Sbe2	0.0009	0.054	0.30	1.799	2.538	2	46	814	89947	124
41	AGL315W	Thp3, Mni2	0.47	10.32	148.65	1.338	2.497	6	33	569	64357	220
42	AFR536W	NOHBY650	0.0003	0.037	0.10	2.085	2.489	1	115	1259	141261	36
43	AER078W	YLR407W	0.0039	0.55	1.17	2.151	2.478	1	26	217	24255	65
44	AEL009C	Med2	0.0007	0.005	0.22	0.798	2.470	1	31	343	37046	56
45	ABL147W	Swt21	0.0003	0.010	0.072	1.581	2.450	2	39	407	45466	96
46	AGR207C	Swd3	0.0021	0.20	0.54	1.979	2.419	1	22	320	34767	58
47	AFR204W	Hsk3	0.0016	0.025	0.41	1.190	2.408	1	7	68	7948	38
48	ACL197W	Prx1	0.0005	0.084	0.12	2.247	2.397	1	21	226	25626	43
49	ADR116C	Vps60	0.011	0.077	2.67	0.832	2.370	3	13	261	28896	200
50	AER341W	Mod5	0.027	1.96	6.33	1.856	2.365	2	26	442	51064	80

Table5-6: SGD based description of the 50 highest up regulated proteins from zero to eleven hours.

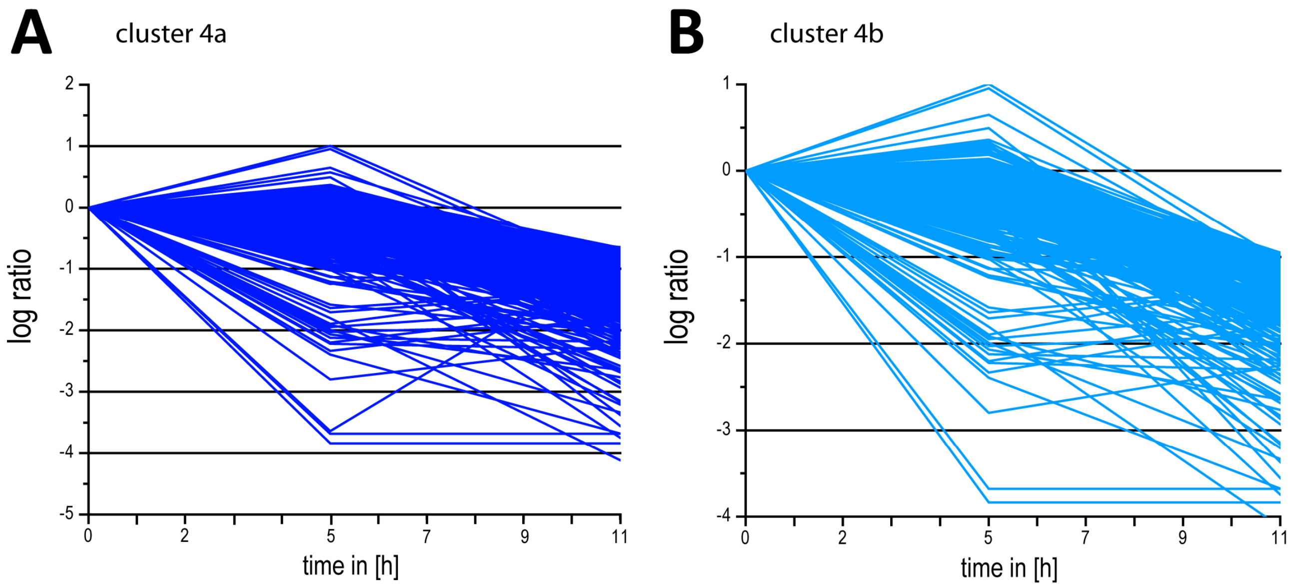
no.	common name	info
1	Zrg8	Protein of unknown function; authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies; GFP-fusion protein is localized to the cytoplasm; transcription induced under conditions of zinc deficiency
2	Ecm5	Non-essential protein of unknown function, contains ATP/GTP-binding site motif A; null mutant exhibits cellular volume up to four times greater than wild-type, also large drooping buds with elongated necks
3	Spc25	Component of the evolutionarily conserved kinetochore-associated Ndc80 complex (Ndc80p-Nuf2p-Spc24p-Spc25p); involved in chromosome segregation, spindle checkpoint activity and kinetochore clustering
4		
5	Snf6	Subunit of the SWI/SNF chromatin remodeling complex involved in transcriptional regulation; functions interdependently in transcriptional activation with Snf2p and Snf5p
6	Cwc15	Non-essential protein involved in pre-mRNA splicing, component of a complex containing Cef1p; has similarity to S. pombe Cwf15p
7	Hap2	Subunit of the heme-activated, glucose-repressed Hap2p/3p/4p/5p CCAAT-binding complex, a transcriptional activator and global regulator of respiratory gene expression; contains sequences sufficient for both complex assembly and DNA binding
8	Mus81	Subunit of the structure-specific Mms4p-Mus81p endonuclease that cleaves branched DNA; involved in DNA repair, replication fork stability, and joint molecule formation/resolution during meiotic recombination; helix-hairpin-helix protein
9	Ccw12	Cell wall mannoprotein, mutants are defective in mating and agglutination, expression is downregulated by alpha-factor
10	Fmp26	Protein of unknown function; the authentic, non-tagged protein is detected in purified mitochondria in high-throughput studies; null mutant displays reduced respiratory growth and elevated frequency of mitochondrial genome loss
11	Mms4	Subunit of the structure-specific Mms4p-Mus81p endonuclease that cleaves branched DNA; involved in recombination, DNA repair, and joint molecule formation/resolution during meiotic recombination
12	Spp381	mRNA splicing factor, component of U4/U6.U5 tri-snRNP; interacts genetically and physically with Prp38p
13	Cse4	Centromere protein that resembles histone H3, required for proper kinetochore function; homolog of human CENP-A; levels are regulated by E3 ubiquitin ligase Psh1p
14	Pex21, Pex18	Peroxin required for targeting of peroxisomal matrix proteins containing PTS2; interacts with Pex7p; partially redundant with Pex18p
15	Nsl1	Essential component of the MIND kinetochore complex (Mtw1p Including Nnf1p-Nsl1p-Dsn1p) which joins kinetochore subunits contacting DNA to those contacting microtubules; required for accurate chromosome segregation
16	Cin5, Yap6	Basic leucine zipper (bZIP) transcription factor of the yAP-1 family, mediates pleiotropic drug resistance and salt tolerance; nuclearly localized under oxidative stress and sequestered in the cytoplasm by Lot6p under reducing conditions
17	Psy2	Putative subunit of an evolutionarily conserved protein phosphatase complex containing the catalytic subunit Pph3p and the regulatory subunit Psy4p; required for cisplatin and oxaliplatin resistance; putative homolog of mammalian R3
18	Mak3	Catalytic subunit of N-terminal acetyltransferase of the NatC type; required for replication of dsRNA virus
19	Bas1	Myb-related transcription factor involved in regulating basal and induced expression of genes of the purine and histidine biosynthesis pathways; also involved in regulation of meiotic recombination at specific genes
20	Snu23	Component of U4/U6.U5 snRNP involved in mRNA splicing via spliceosome
21	Dsn1	Essential component of the MIND kinetochore complex (Mtw1p Including Nnf1p-Nsl1p-Dsn1p) which joins kinetochore subunits contacting DNA to those contacting microtubules; important for chromosome segregation
22	Duo1	Essential subunit of the Dam1 complex (aka DASH complex), couples kinetochores to the force produced by MT depolymerization thereby aiding in chromosome segregation; is transferred to the kinetochore prior to mitosis
23	YKL075C	Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm; proposed to be involved in resistance to streptozotocin and camptothecin
24	YOL036W, YIR01	Protein of unknown function; potential Cdc28p substrate
25	Smf3	Putative divalent metal ion transporter involved in iron homeostasis; transcriptionally regulated by metal ions; member of the Nramp family of metal transport proteins
26	Slx9	Protein required for pre-rRNA processing; associated with the 90S pre-ribosome and 43S small ribosomal subunit precursor; interacts with U3 snoRNA; deletion mutant has synthetic fitness defect with an sgs1 deletion mutant
27	Sme1	Core Sm protein Sm E; part of heteroheptameric complex (with Smb1p, Smd1p, Smd2p, Smd3p, Smx3p, and Smx2p) that is part of the spliceosomal U1, U2, U4, and U5 snRNPs; homolog of human Sm E
28	Ogg1	Mitochondrial glycosylase/lyase that specifically excises 7,8-dihydro-8-oxoguanine residues located opposite cytosine or thymine residues in DNA, repairs oxidative damage to mitochondrial DNA, contributes to UVA resistance
29	Ndl1	Homolog of nuclear distribution factor NudE, NUDEL; interacts with Pac1p and regulates dynein targeting to microtubule plus ends
30	Rec8	Meiosis-specific component of sister chromatid cohesion complex; maintains cohesion between sister chromatids during meiosis I; maintains cohesion between centromeres of sister chromatids until meiosis II; homolog of S. pombe Rec8p
31	YHR132W-A, YNL	Protein required for initiation of G0 program; prevents degradation of nutrient-regulated mRNAs via the 5'-3' mRNA decay pathway; phosphorylated by Rim15p; GFP protein localizes to the cytoplasm and nucleus; similar to Igo1p
32	Dot6	Protein involved in rRNA and ribosome biogenesis; binds polymerase A and C motif; subunit of the RPD3L histone deacetylase complex; similar to Tod6p; has chromatin specific SANT domain; involved in telomeric gene silencing and filamentation
33	Nnf1	Essential component of the MIND kinetochore complex (Mtw1p Including Nnf1p-Nsl1p-Dsn1p) which joins kinetochore subunits contacting DNA to those contacting microtubules; required for accurate chromosome segregation
34	Vph2	Integral membrane protein required for vacuolar H+-ATPase (V-ATPase) function, although not an actual component of the V-ATPase complex; functions in the assembly of the V-ATPase; localized to the endoplasmic reticulum (ER)
35	Shg1	Subunit of the COMPASS (Set1C) complex, which methylates histone H3 on lysine 4 and is required in transcriptional silencing near telomeres
36	Pet191	Protein required for assembly of cytochrome c oxidase; exists as an oligomer that is integral to the mitochondrial inner membrane and faces the intermembrane space; contains a twin Cx9C motif
37	Nut2	Subunit of the RNA polymerase II mediator complex; associates with core polymerase subunits to form the RNA polymerase II holoenzyme; required for transcriptional activation and has a role in basal transcription
38	Irc8	Bud tip localized protein of unknown function; mRNA is targeted to the bud by a She2p dependent transport system; mRNA is cell cycle regulated via Fkh2p, peaking in G2/M phase; null mutant displays increased levels of spontaneous Rad52p foc
39	Mga1	Protein similar to heat shock transcription factor; multicopy suppressor of pseudohyphal growth defects of ammonium permease mutants
40	Sbe22, Sbe2	Protein involved in the transport of cell wall components from the Golgi to the cell surface; similar in structure and functionally redundant with Sbe2p; involved in bud growth
41	Thp3, Mni2	Protein that forms a complex with Csn12p that is recruited to transcribed genes and may have a role in transcription elongation; possibly involved in splicing based on pre-mRNA accumulation defect for many intron-containing genes
42	NOHBY650	
43	YLR407W	Putative protein of unknown function; null mutant displays elongated buds and a large fraction of budded cells have only one nucleus
44	Med2	Subunit of the RNA polymerase II mediator complex; associates with core polymerase subunits to form the RNA polymerase II holoenzyme; essential for transcriptional regulation
45	Swt21	Protein involved in mRNA splicing; contains a consensus nuclear export signal (NES) sequence similar to the consensus sequence recognized by Crm1p; interacts genetically with Prp40p and Tgs1p; contains WD40 repeats
46	Swd3	Essential subunit of the COMPASS (Set1C) complex, which methylates histone H3 on lysine 4 and is required in transcriptional silencing near telomeres; WD40 beta propeller superfamily member and ortholog of mammalian WDR5
47	Hsk3	Essential subunit of the Dam1 complex (aka DASH complex), couples kinetochores to the force produced by MT depolymerization thereby aiding in chromosome segregation; is transferred to the kinetochore prior to mitosis
48	Prx1	Mitochondrial peroxiredoxin (1-Cys Prx) with thioredoxin peroxidase activity, has a role in reduction of hydroperoxides; reactivation requires Trr2p and glutathione; induced during respiratory growth and oxidative stress; phosphorylated
49	Vps60	Cytoplasmic and vacuolar membrane protein involved in late endosome to vacuole transport; required for normal filament maturation during pseudohyphal growth; may function in targeting cargo proteins for degradation; interacts with Vta1p
50	Mod5	Delta 2-isopentenyl pyrophosphate:tRNA isopentenyl transferase, required for biosynthesis of the modified base isopentenyladenosine in mitochondrial and cytoplasmic tRNAs; gene is nuclear and encodes two isozymic forms

Table5-7: Ranking of the 50 most down regulated proteins from zero to eleven hours.

no.	accession	common name	0h relative abundance (all *1)	5h relative abundance (all *1)	11h relative abundance (all *1)	5h log ratio	11h log ratio	detected peptieds	possible peptides	protein size [aa]	protein size [MW]	Score
1	AGR359C	Ppe1	5.26	0.082	0.00040	-1.805	-4.119	1	25	385	42796	47
2	ACL125C	Lsm8	17.15	0.0025	0.0025	-3.836	-3.836	1	4	106	11421	56
3	ACL142W	Gpi1	1.55	0.33	0.00028	-0.669	-3.747	3	36	661	75864	104
4	AFR522C	Ste2	1.78	0.00037	0.00037	-3.681	-3.681	1	27	458	50476	68
5	AGR225W		11.93	0.048	0.0025	-2.395	-3.679	1	4	94	10083	50
6	AER072W	Fld1, Seipin|Sei1	2.24	2.91	0.00063	0.114	-3.554	1	16	261	29564	53
7	ADL126C	Rrt5	0.49	0.55	0.00021	0.053	-3.372	2	48	466	52276	53
8	AAL004W	Ldb16	0.80	0.0082	0.00037	-1.990	-3.334	2	27	311	35020	55
9	AER125C	Lsm3	10.78	1.03	0.0067	-1.019	-3.206	1	10	83	9167	54
10	ABR053C	Erg5	0.28	0.13	0.00019	-0.327	-3.166	1	52	515	58702	46
11	AFR036C	YHR140W	44.55	13.40	0.032	-0.522	-3.147	1	11	224	24295	51
12	ADL138C	Die2	2.28	0.83	0.0027	-0.437	-2.931	2	24	509	59061	112
13	AGL126C	YLR168C, YDR185C	0.29	0.20	0.00040	-0.157	-2.866	1	25	222	25555	37
14	AEL199W	Shr5	4.09	0.82	0.0060	-0.696	-2.831	2	12	237	27213	124
15	AMI008W	Cob	5.19	0.063	0.0089	-1.916	-2.764	1	7	384	43802	36
16	AFR508W	YLR050C	7.28	1.40	0.015	-0.715	-2.683	3	8	167	19829	127
17	AGL002C	Snf8	0.29	0.0025	0.00067	-2.069	-2.644	1	15	232	26842	39
18	AFR604C	Spo21	0.13	0.014	0.00031	-0.994	-2.643	2	56	844	96461	103
19	AFR090W	Phs1	3.41	0.71	0.0081	-0.683	-2.627	2	12	217	24194	102
20	AGR099C	Arp10	0.30	0.061	0.00079	-0.693	-2.580	1	31	305	34133	66
21	AGR263C	Nej1	2.99	0.94	0.011	-0.502	-2.452	2	30	270	31433	89
22	AAR175C	Mon1	2.40	3.01	0.0093	0.099	-2.412	2	25	581	64659	65
23	AGL071C	Mot3	1.93	0.31	0.0076	-0.795	-2.404	2	8	396	20056	117
24	AFR002C	Vps8	0.032	0.028	0.00014	-0.047	-2.364	1	73	1230	139423	39
25	AGL011C	NOHBY701	5.95	0.72	0.026	-0.916	-2.352	1	17	317	34779	64
26	AGL365C	NOHBY725	0.11	0.00064	0.00053	-2.221	-2.301	1	46	599	67207	44
27	ABR223C	YFL040W	3.42	2.22	0.017	-0.187	-2.296	2	28	624	68959	141
28	AGR408W		0.78	0.0013	0.0044	-2.798	-2.252	1	8	172	19278	78
29	AGR326C	YJR011C	4.94	0.37	0.028	-1.126	-2.241	1	10	202	23132	37
30	AFR415C	Nta1	9.34	3.62	0.057	-0.412	-2.215	4	28	416	47678	190
31	ABR166C	Ilm1	24.86	12.75	0.15	-0.290	-2.215	2	9	177	19970	94
32	AAL184W	YML108W	1.71	0.013	0.011	-2.123	-2.186	1	9	112	12810	35
33	AFR711C	Fig2	0.90	1.11	0.0059	0.093	-2.180	1	24	561	57579	63
34	ADL292C	YDL114W	1.56	1.75	0.010	0.052	-2.175	4	21	329	36594	109
35	ADR119W	YBR284W, YJL070C	1.16	0.21	0.0085	-0.746	-2.139	4	41	748	87639	106
36	ACL154W	Hmx1	0.13	0.42	0.0010	0.494	-2.118	1	37	319	35394	42
37	AFL013C		1.00	0.55	0.0078	-0.262	-2.108	1	14	216	24401	44
38	AGR337W	Got1	3.68	1.01	0.030	-0.560	-2.081	1	7	138	15148	74
39	AEL125C	Tpo5	0.037	0.039	0.00031	0.022	-2.075	1	32	575	62442	44
40	AFR523C	Tim18	170.47	70.41	1.46	-0.384	-2.067	3	5	180	20112	71
41	AEL262C		147.73	38.71	1.29	-0.582	-2.059	1	5	85	9210	49
42	AAL089W	YOR246C	23.17	0.61	0.21	-1.583	-2.036	3	14	328	37370	136
43	AER287W	Mdm1	0.038	0.0023	0.00037	-1.227	-2.013	2	82	1153	133676	83
44	AAR104W	Cog2	58.78	23.76	0.59	-0.393	-1.999	6	15	255	28973	317
45	ABR088C	Elm1	0.35	0.42	0.0036	0.080	-1.985	1	56	532	60031	39
46	AER010C	Dsf2	0.013	0.0075	0.00013	-0.227	-1.977	1	75	764	84536	42
47	AER096C	Cnb1	115.53	6.68	1.29	-1.238	-1.951	4	12	175	19713	213
48	AER005C	Fmp44	4.95	2.54	0.059	-0.290	-1.924	3	20	427	49991	110
49	AAL105C	Osw1	6.95	6.61	0.084	-0.022	-1.917	5	16	213	24382	220
50	ACR156W	Rny1	2.17	0.020	0.027	-2.027	-1.906	3	20	292	33394	92

Table5-8: SGD based description of the 50 highest down regulated proteins from zero to eleven hours.

no.	common name	info
1	Ppe1	Protein with carboxyl methyl esterase activity that may have a role in demethylation of the phosphoprotein phosphatase catalytic subunit; also identified as a small subunit mitochondrial ribosomal protein
2	Lsm8	Lsm (Like Sm) protein; forms heteroheptameric complex (with Lsm2p, Lsm3p, Lsm4p, Lsm5p, Lsm6p, and Lsm7p) that is part of spliceosomal U6 snRNP and is also implicated in processing of pre-tRNA, pre-snoRNA, and pre-rRNA
3	Gpi1	Membrane protein involved in the synthesis of N-acetylglucosaminyl phosphatidylinositol (GlcNAc-PI), the first intermediate in the synthesis of glycosylphosphatidylinositol (GPI) anchors; human and mouse GPI1p are functional homologs
4	Ste2	Receptor for alpha-factor pheromone; seven transmembrane-domain GPCR that interacts with both pheromone and a heterotrimeric G protein to initiate the signaling response that leads to mating between haploid a and alpha cells
5	AGR225W	
6	Fld1, Seipin|Sei1	Seipin protein involved in lipid droplet morphology, number, and size; proposed to be involved in lipid metabolism; related to the human BSCL2 which is associated with lipodystrophy
7	Rrt5	Putative protein of unknown function; non-essential gene identified in a screen for mutants with increased levels of rDNA transcription; expressed at high levels during sporulation
8	Ldb16	Protein of unknown function; null mutants have decreased net negative cell surface charge; GFP-fusion protein expression is induced in response to the DNA-damaging agent MMS; native protein is detected in purified mitochondria
9	Lsm3	Lsm (Like Sm) protein; part of heteroheptameric complexes (Lsm2p-7p and either Lsm1p or 8p): cytoplasmic Lsm1p complex involved in mRNA decay; nuclear Lsm8p complex part of U6 snRNP and possibly involved in processing tRNA, snoRNA, and rRNA
10	Erg5	C-22 sterol desaturase, a cytochrome P450 enzyme that catalyzes the formation of the C-22(23) double bond in the sterol side chain in ergosterol biosynthesis; may be a target of azole antifungal drugs
11	YHR140W	Putative integral membrane protein of unknown function
12	Die2	Dolichyl-phosphoglucose-dependent alpha-1,2 glucosyltransferase of the ER, functions in the pathway that synthesizes the dolichol-linked oligosaccharide precursor for N-linked protein glycosylation, has a role in regulation of ITR1 and INO1
13	YLR168C, YDR185C	Mitochondrial intermembrane space protein involved in regulation of mitochondrial cardiolipin and phosphatidylethanolamine levels; null has defects in mitochondrial morphology; similar to Ups1p, Ups3p and to human PRELI
14	Shr5	Subunit of a palmitoyltransferase, composed of Shr5p and Erf2p, that adds a palmitoyl lipid moiety to heterolipidated substrates such as Ras1p and Ras2p through a thioester linkage; palmitoylation is required for Ras2p membrane localization
15	Cob	
16	YLR050C	Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the endoplasmic reticulum; YLR050C is not an essential
17	Snf8	Component of the ESCRT-II complex, which is involved in ubiquitin-dependent sorting of proteins into the endosome; appears to be functionally related to SNF7; involved in glucose derepression
18	Spo21	Component of the meiotic outer plaque of the spindle pole body, involved in modifying the meiotic outer plaque that is required prior to prospore
19	Phs1	Essential 3-hydroxyacyl-CoA dehydratase of the ER membrane, involved in elongation of very long-chain fatty acids; evolutionarily conserved, similar to mammalian PTPLA and PTPLB; involved in sphingolipid biosynthesis and protein trafficking
20	Arp10	Component of the dynactin complex, localized to the pointed end of the Arp1p filament; may regulate membrane association of the complex
21	Nej1	Protein involved in regulation of nonhomologous end joining; interacts with DNA ligase IV components Dnl4p and Lif1p; repressed by MAT heterozygosity; regulates cellular distribution of Lif1p
22	Mon1	Protein required for fusion of cvt-vesicles and autophagosomes with the vacuole; associates, as a complex with Ccz1p, with a perivacuolar compartment; potential Cdc28p substrate
23	Mot3	Nuclear transcription factor with two Cys2-His2 zinc fingers; involved in repression of a subset of hypoxic genes by Rox1p, repression of several DAN/TIR genes during aerobic growth, and repression of ergosterol biosynthetic genes
24	Vps8	Membrane-associated protein that interacts with Vps21p to facilitate soluble vacuolar protein localization; component of the CORVET complex; required for localization and trafficking of the CPY sorting receptor; contains RING finger motif
25	NOHBY701	
26	NOHBY725	
27	YFL040W	Putative transporter, member of the sugar porter family; YFL040W is not an essential gene
28	AGR408W	
29	YJR011C	Putative protein of unknown function; GFP-fusion protein expression is induced in response to the DNA-damaging agent MMS
30	Nta1	Amidase, removes the amide group from N-terminal asparagine and glutamine residues to generate proteins with N-terminal aspartate and glutamate residues that are targets of ubiquitin-mediated degradation
31	Ilm1	Protein of unknown function; may be involved in mitochondrial DNA maintenance; required for slowed DNA synthesis-induced filamentous growth
32	YML108W	Putative protein of unknown function whose structure defines a new subfamily of the split beta-alpha-beta sandwiches; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm and nucleus; YML108W is not an essential gene
33	Fig2	Cell wall adhesin, expressed specifically during mating; may be involved in maintenance of cell wall integrity during mating
34	YDL114W	Putative protein of unknown function with similarity to acyl-carrier-protein reductases; YDL114W is not an essential gene
35	YBR284W, YJL070C	Putative protein of unknown function; YBR284W is not an essential gene; null mutant exhibits decreased resistance to rapamycin and wortmannin and synthetic phenotype with alpha-synuclein
36	Hmx1	ER localized heme oxygenase, involved in heme degradation during iron starvation and in the oxidative stress response; expression is regulated by AFT1 and oxidative stress; relocates to the perinuclear region in the presence of oxidants
37	AFL013C	
38	Got1	Homodimeric protein that is packaged into COPII vesicles and cycles between the ER and Golgi; involved in secretory transport but not directly required for aspects of transport assayed in vitro; may influence membrane composition
39	Tpo5	Protein involved in excretion of putrescine and spermidine; putative polyamine transporter in the Golgi or post-Golgi vesicles
40	Tim18	Component of the mitochondrial TIM22 complex involved in insertion of polytopic proteins into the inner membrane; may mediate assembly or stability of
41	AEL262C	
42	YOR246C	Protein with similarity to oxidoreductases, found in lipid particles; required for replication of Brome mosaic virus in S. cerevisiae, a model system for studying replication of positive-strand RNA viruses in their natural hosts
43	Mdm1	Intermediate filament protein, required for nuclear and mitochondrial transmission to daughter buds; contains a Phox homology (PX) domain and specifically binds phosphatidylinositol 3-phosphate (PtdIns-3-P)
44	Cog2	Essential component of the conserved oligomeric Golgi complex (Cog1p through Cog8p), a cytosolic tethering complex that functions in protein trafficking to mediate fusion of transport vesicles to Golgi compartments
45	Elm1	Serine/threonine protein kinase that regulates cellular morphogenesis, septin behavior, and cytokinesis; required for the regulation of other kinases; forms part of the bud neck ring
46	Dsf2	Deletion suppressor of mpt5 mutation
47	Cnb1	Calcineurin B; the regulatory subunit of calcineurin, a Ca++/calmodulin-regulated type 2B protein phosphatase which regulates Crz1p (a stress-response transcription factor), the other calcineurin subunit is encoded by CNA1 and/or CMP1
48	Fmp44	Functional ortholog of human PIG-V, which is a mannosyltransferase that transfers the second mannose in glycosylphosphatidylinositol biosynthesis; the authentic, non-tagged protein was localized to mitochondria
49	Osw1	Protein involved in sporulation; required for the construction of the outer spore wall layers; required for proper localization of Spo14p
50	Rny1	Vacuolar RNase of the T(2) family, relocalizes to the cytosol where it cleaves tRNAs upon oxidative or stationary phase stress; promotes apoptosis under stress conditions and this function is independent of its catalytic activity



function - cluster 4a	**count**	**enrichment score**	**P-Value**
intrinsic to membrane	88	4.89	0.000000049
transmembrane	64	4.89	0.0016
proteolysis	41	2.54	0.00038
endoplasmic reticulum	28	2.09	0.0011
Starch and sucrose metabolism	14	1.92	0.000026
autophagy	13	1.48	0.0025
energy derivation by oxidation of organic compounds	12	1.3	0.0033
exopeptidase activity	7	1.26	0.0032
septin cytoskeleton	5	0.8	0.028
total	**866**		

function - 4b	**count**	**enrichment score**	**P-Value**
integral to membrane	48	2.96	8.80E-06
transmembrane	36	2.96	8.70E-03
glycoprotein	10	2.39	0.0094
Starch and sucrose metabolism	9	1.69	0.00055
iron ion binding	12	1.26	0.010
phosphate metabolic process	20	0.61	0.048
proteolysis	15	0.16	0.280
total	**456**		

Figure5-9: All down regulated proteins in eleven hour young mycelium. **(A)** Minor decrease in the protein abundance with the list of biological reduced protein groups in this cluster. **(B)** Only proteins that are decreased by a major value including the functional annotation of the down regulated protein groups.

The group of proteins which are degraded during the development from spores to young mycelium was manually annotated. In total we could identify over 800 proteins are most likely no longer present in germ bubbles and / or young mycelia. The top 50 decreased proteins at eleven hours are presented in Table5-7 and their cellular functions in Table5-8. To gain a better understanding of these proteins we analyzed the cluster 4a and b with the DAVID tool (Figure5-9). This lead to a global view of the most significantly changed functional protein categories. Some of the identified groups were basically already identified at five hours, but not with a lower protein counts and a higher P-values. Only 41 proteins were degraded, which indicates that it is less essential to degrade proteins than to express huge amounts of novel proteins. Also autophagy proteins and the exopeptidase are down-regulated during the development from spores to young mycelium. Suggesting that those biological functions play a higher role during sporulation, the presence of protein categories in this cluster is not surprising. With the incubation in high glucose liquid full medium the observed down-regulation of proteins belonging to the starch and sucrose metabolism can be explained. With a P-value of 0.000026 the significance for these pathways to be down-regulated is evident. 64 transmembrane proteins with a low P-value were also identified. Furthermore, 88 membrane proteins are found in this down-regulated cluster. The global group of proteins that carry moieties, which are inserted in the membrane such as a GPI anchor is not necessarily actively down-regulated. Because of the increase of the membrane surrounding volume the percentage of the volume expands faster than the bilayer during the germination from spores to young mycelium. Compared to the cluster 4a the cluster 4b with the more than 10 times decreasing protein abundances carries the same functional categories, only a more specific protein group is significantly enriched. Ten glycoproteins with a P-value of 0.0094 were identified. This group of proteins is mainly located at cell membranes and is probably reduced in their abundances because of the same reason (volume, perimeter ratio) as the other membrane proteins.

K-means clustering of protein species during the development of *A. gossypii*

Development is characterized by dramatic variations in the protein levels over time. We did a K-means clustering using the ratio of logarithmic scales of protein abundances. Figure5-10 shows the clustering results and the developing proteins species during germination of *A. gossypiiI*. The theoretical description of the behavior of the log values over the time is presented by an overview graphic (Figure5-10A). Included are the limitations of the log values for the different clusters as well as the specification for the five and eleven hours time points. In total we obtain 5 clusters with general protein characterization. Cluster 1 represents proteins that are induced and cluster 2 those proteins that are decreasing. Cluster 3 contains proteins that are first up-regulated at five hours and down- or un-regulated after eleven hours. Cluster 4 shows proteins that are first down-regulated during the first five hours and are than increasing during later stages. Cluster 5 covers the proteins that are mainly un-regulated during the first eleven hours of the development of A. gossypii. For a more specific interpretation we split the up- and down-regulated clusters into subassembly groups of minor and major alterations plus changes that only occur after five hours. In Figure5-10B the average logarithmic ratio of the five clusters including the sub clusters is plotted. The general behavior of the proteins that are grouped in one cluster are easily comparable. As example, the up-regulated sub clusters are more like the theoretical predictions compared with the down-regulate sub clusters, which all show a more or less similar curve. In contrast, the induced clusters show specific log ratio differences at the five hour time point. The total number of proteins per cluster is shown in the chart diagram (Figure5-10C). The largest and most prominent cluster with over 2,000 proteins is the unregulated cluster 5. There are 800 up- and also about 800 down-regulated proteins. For only about 100 proteins we up-regulation followed by down-regulation and vice versa.
The first cluster with all proteins that share the same characterization of increased abundance over the time points is presented in Figure5-11. The logarithm to the base 10 protein abundances of all cluster 1 proteins are plotted together (Figure5-11A). The subassembly groups of this cluster are divided into three clusters 1a, 1b, and 1c (Figure5-11B). In cluster 1a all proteins are grouped that show a more than 10-fold increase after elven hours. Proteins that only change by a factor of five to ten after eleven hours are assembled in cluster 1b. The last cluster contains up-regulated proteins that are stable during the first five hours and then become induced. The functional categories that

are associated with cluster 1 proteins and the sub-groups are listed in Table5-9. As described previously for the time point clusters (Figure5-6&8) we could identify similar functional groups.

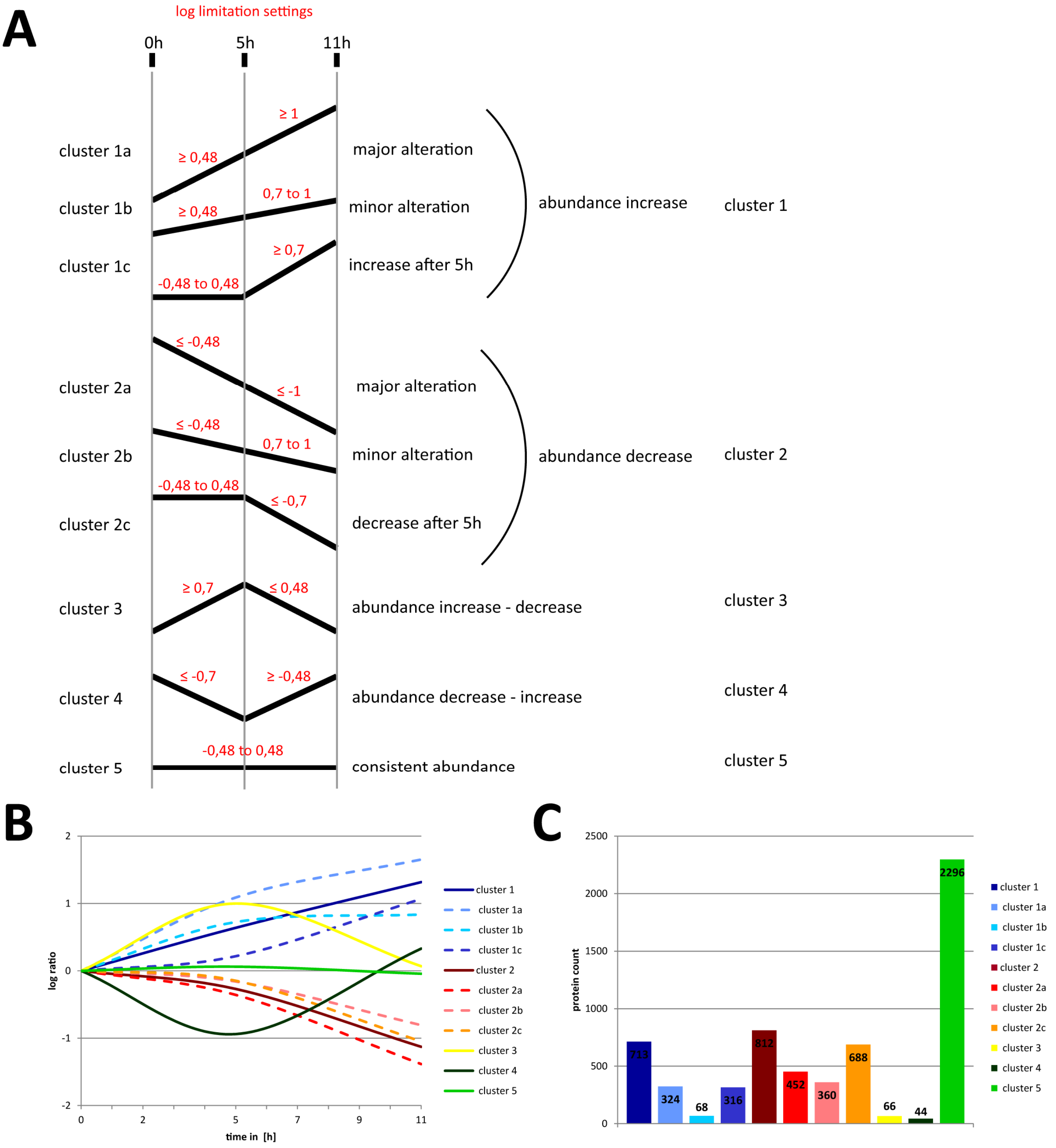


Figure5-10: K-means clustering of protein abundances during the development of *A. gossypii*. **(A)** Theoretical cluster description with the log values changes and the limitation settings for the different time points. **(B)** Average of each cluster over the time points. **(C)** Protein species numbers of the different clusters.

Especially proteins for ribosome biogenesis, regulation of transcription, RNA processing, and RNA-binding are significantly enriched in the cluster of up-regulated proteins. Furthermore, proteins involved in cell division and DNA replication are found to be substantially induced. In the sub-clusters 1a we found many expected protein groups based on Figure5-8. In the function cluster 1a of Table5-9 proteins are found involved in cell growth, like biogenesis for ribonucleoproteins or ten sub-units of the RNA polymerase. Among proteins that are moderately induced we found for example proteins of non-membrane-bound organelles, in a cluster of only 68 proteins. Interestingly, we found the same protein categories in cluster 1a and 1c, like the ribonucleoprotein complex proteins, which suggests that these cellular functions become more and more important at later stages of the life cycle of *A. gossypii* to achieve the huge increases of biomass and levels of proteins.

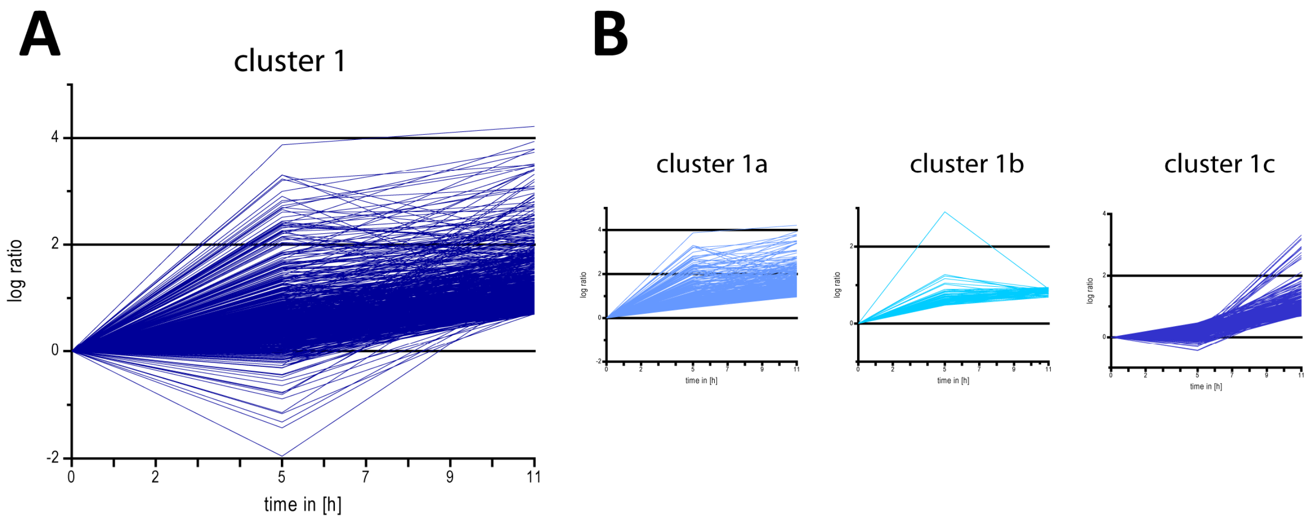


Figure5-11: K-means cluster of up-regulated proteins. **(A)** All up-regulated proteins combined in cluster 1. **(B)** Separation of cluster 1 into three sub-groups with different specification of up-regulation.

Table5-9: Functional annotated protein groups based on the DAVID Bioinformatics tool.

function - cluster 1	count	enrichment score	P-Value
ribosome biogenesis	56	18.13	5.7E-27
RNA processing	68	18.13	2.8E-13
non-membrane-bounded organelle	114	9.74	1.2E-24
RNA recognition motif, RNP-1	20	3.65	0.00005
regulation of transcription	64	3.36	0.000001
rna-binding	24	1.4	0.000024
cell division	13	1.24	0.0044
chromosome organization	18	1.17	0.032
DNA replication	12	0.97	0.02
metal ion binding	68	0.56	0.72
total	**713**		

function - cluster 1a	count	enrichment score	P-Value
ribonucleoprotein complex biogenesis	37	15.28	7.8E-22
Pyrimidine metabolism	16	4.16	0.000000072
RNA polymerase	10	4.16	0.0000038
Purine metabolism	14	4.16	0.000036
regulation of transcription	28	1.61	0.0044
cell division	7	1.21	0.032
rna-binding	13	0.78	0.00099
zinc ion binding	30	0.74	0.0095
total	**325**		

function - cluster1b	count	enrichment score	P-Value
non-membrane-bounded organelle	11	1.61	0.0082
chromosome	5	1.61	0.029
ribonucleoprotein complex	6	1.61	0.18
mitochondrial membrane	3	0.23	0.44
total	**68**		

function - cluster1c	count	enrichment score	P-Value
ribonucleoprotein complex	38	7.93	4.5E-09
Ribosome	24	7.93	2E-12
RNA processing	22	1.95	0.0011
regulation of transcription	31	1.68	0.00039
sequence-specific DNA binding	9	1.83	0.0019
prefoldin complex	4	1.1	0.013
metal-binding	19	0.24	0.15
RNA polymerase	5	0.87	0.06
total	**316**		

Over 800 proteins were found to be down-regulated. In Figure5-12 the K-means cluster is shown of the down-regulated proteins (Figure5-12A) and of the subassembly groups (Figure5-12B). The corresponding functional annotations of the cluster are listed in Table5-10. In general we could identify the same protein families like as in the time point clustering. The intrinsic membrane annotation was detected in all three sub-clusters with over 80 proteins. Similarly, peptidases other degrading enzymes are all enriched in the three sub-clusters 2a-c. Interestingly, we could detect a significant decrease of 14 starch and sucrose metabolism proteins in cluster 1c, which indicates either active degradation inhibition of gene expression. As said before during growth in glucose rich medium these additional metabolic functions, are no longer needed. As calculated from cluster 2c unspecific transport proteins are reduced in general. The P-value of the 43 transport proteins is 0.0026, which reflects a significant enrichment in this sub-cluster.

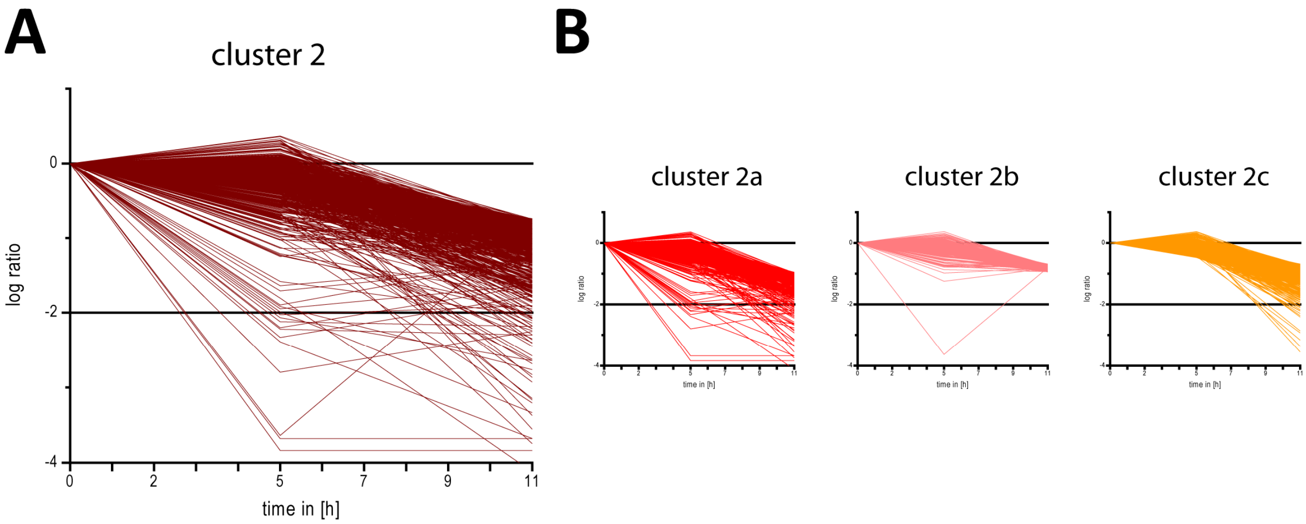


Figure5-12: K-means cluster of the down-regulated proteins. **(A)** All down-regulated proteins combined in cluster 2. **(B)** Separation of cluster 2 into three subassembly groups with different specification of down regulation.

Table5-10: Functional annotated protein groups based on the DAVID Bioinformatics tool of the cluster 2.

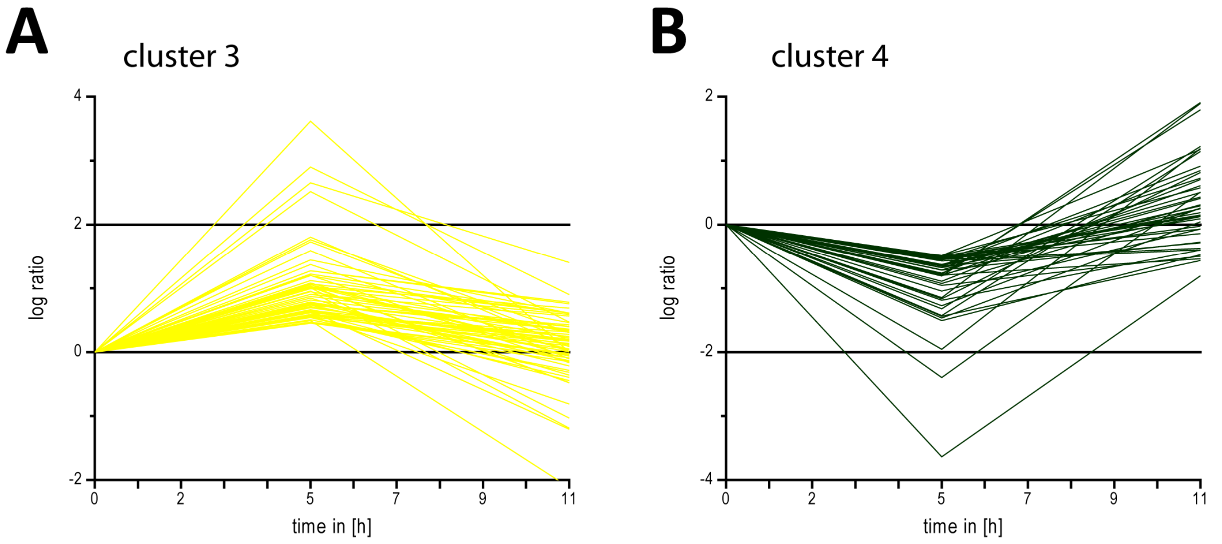
function - cluster 2	**count**	**enrichment score**	**P-Value**
intrinsic to membrane	86	5.31	0.000000023
endoplasmic reticulum	28	2.4	0.00056
peptidase activity	29	2.24	0.00041
proteolysis	37	2.24	0.0015
Starch and sucrose metabolism	14	2.24	0.000012
glycosylation site:N-linked (GlcNAc...)	13	2.01	0.0059
transport	51	1.67	0.00071
generation of precursor metabolites and energy	23	1.57	0.0024
exopeptidase activity	7	1.25	0.0021
carboxylesterase activity	7	1.4	0.02
total	**812**		

function - cluster 2a	**count**	**enrichment score**	**P-Value**
integral to membrane	48	2.99	8.8E-06
Starch and sucrose metabolism	7	1.71	0.00055
generation of precursor metabolites and energy	13	1.01	0.017
iron ion binding	12	1.28	0.009
proteolysis	14	0.87	0.36
exopeptidase activity	5	0.87	0.0086
total	**451**		

function - cluster 2b	**count**	**enrichment score**	**P-Value**
peptidase activity	18	2.46	0.00018
intrinsic to membrane	38	1.97	0.0019
Glycolysis / Gluconeogenesis	6	1.38	0.093
autophagy	7	1.28	0.024
proteasome complex	6	1.08	0.014
regulation of Ras protein signal transduction	6	0.93	0.029
total	**360**		

function - cluster 2c	**count**	**enrichment score**	**P-Value**
intrinsic to membrane	76	4.76	2.3E-07
peptidase activity	27	2.2	0.00017
endopeptidase activity	13	2.2	0.021
Starch and sucrose metabolism	14	2.16	2.5E-06
generation of precursor metabolites and energy	21	1.73	0.0025
transmembrane region	25	1.6	0.013
Alcohol dehydrogenase superfamily, zinc-contain	6	1.54	0.0097
transport	43	1.54	0.0026
proteolysis	33	0.85	0.0022
endoplasmic reticulum	24	1.43	0.0028
total	**688**		

Also we had a look at proteins that undergo different regulations over the time points (Figure5-13), as indicated by different log ratio changes between five and eleven hours. In Figure5-13A cluster 3 is plotted and in the Table the corresponding annotation categories. Only 65 proteins specifically up-regulated in germ bubbles could be identified. DNA binding proteins are significantly enriched with a P-value of 0.026. The eight proteins which are strongly enriched in this cluster show a decrease after eleven hours. Cluster 4 presented in Figure5-13B represents down-regulated proteins at five hours, which are found to be increased after elven hours. With 44 proteins that show this behavior cluster 4 comprises the smallest group of proteins in this analysis. A P-value smaller than 0.05 is only detected by the annotated category of cell division. Considering the P-value of 0.002 this protein group demonstrates a significant enrichment is this cluster. Since cell division proteins play a role during sporulation and at later cell stages as well for septation this annotated group of proteins is biologically meaningful. When the regulation of proteins changes twice during development one can assume that these proteins have a specific function at a short time and are probably needed at certain cell stages. On the other hand such predictions are difficult because the development of *A. gossypii* is a rather asynchronous process.

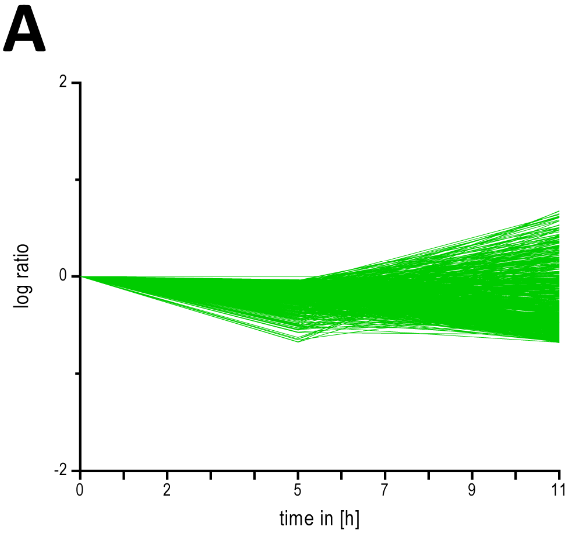


function - cluster 3	count	enrichment score	P-Value
DNA binding	8	0.52	0.026
mitochondrial inner membrane	3	0.45	0.22
ATP binding	5	0.22	0.73
nucleotide binding	8	0.22	0.6
total	**65**		

function - cluster 4	count	enrichment score	P-Value
cell division	4	1.51	0.002
metal ion binding	8	0.84	0.098
total	**44**		

Figure5-13: K-means cluster of proteins that are different regulated at the five and at the eleven hour time points. **(A)** Proteins that are increased at five hours and degraded after eleven hours. **(B)** Down-regulated proteins from zero to five hours that are up-regulation after eleven hours.

The formed cluster of un-regulated proteins is presented in Figure5-14 and included in total 2,295 protein, which is more than the half of the 3,892 identified proteins. Cluster 5 shows the proteins that are stable in their abundance. These proteins have to be actively expressed during germination and formation of young mycelia, because of the huge biomass and protein mass increases during development. With the volume increase from spores to young mycelia the abundance of no longer expressed proteins should be slowly diluted over the time. The functional annotation categories that are significantly enriched in this group are for example; 248 nucleotide-binding proteins and 126 proteins involved in nitrogen metabolism. Mitochondrial proteins are considered to be significantly enriched with a P-value of 0.00074. Other proteins that are expressed during germination and do not significantly further increase are involved in tRNA metabolic processes.



B

function	count	enrichment score	P-Value
nucleotide-binding	248	7.97	3.1E-16
tRNA metabolic process	56	4.95	0.000027
nitrogen compound biosynthetic process	126	3.9	0.00000018
ATPase, AAA+ type, core	54	2.74	0.00088
protein transport	66	2.68	0.0033
intracellular transport	113	2.68	0.00028
mitochondrion	103	2.67	0.00074
endomembrane system	47	1.99	0.00069
GTPase activity	20	1.92	0.00085
Chaperone	29	1.65	0.0025
total	**2295**		

Figure5-14: K-means cluster of proteins that are stable in their abundances during the development of *A. gossypii*. Functional annotation categories of this cluster are listed below.

Cluster of orthologous groups (COGs) based on *S. cerevisiae* annotations

Clusters of Orthologous Groups of proteins (COGs) are characterized by comparing protein species determined in complete genomes (Tatusov et al., 2003). The fraction of protein classifications are calculated by the quantitatively determined protein levels and are shown in percent of the proteome. The quantitative protein levels are calculated via the protein copy numbers using internal heavy peptide standards, which are used to globally estimate absolute protein concentrations for all identified proteins. This analysis predicts fractions of specific biological functions by the total protein masses.

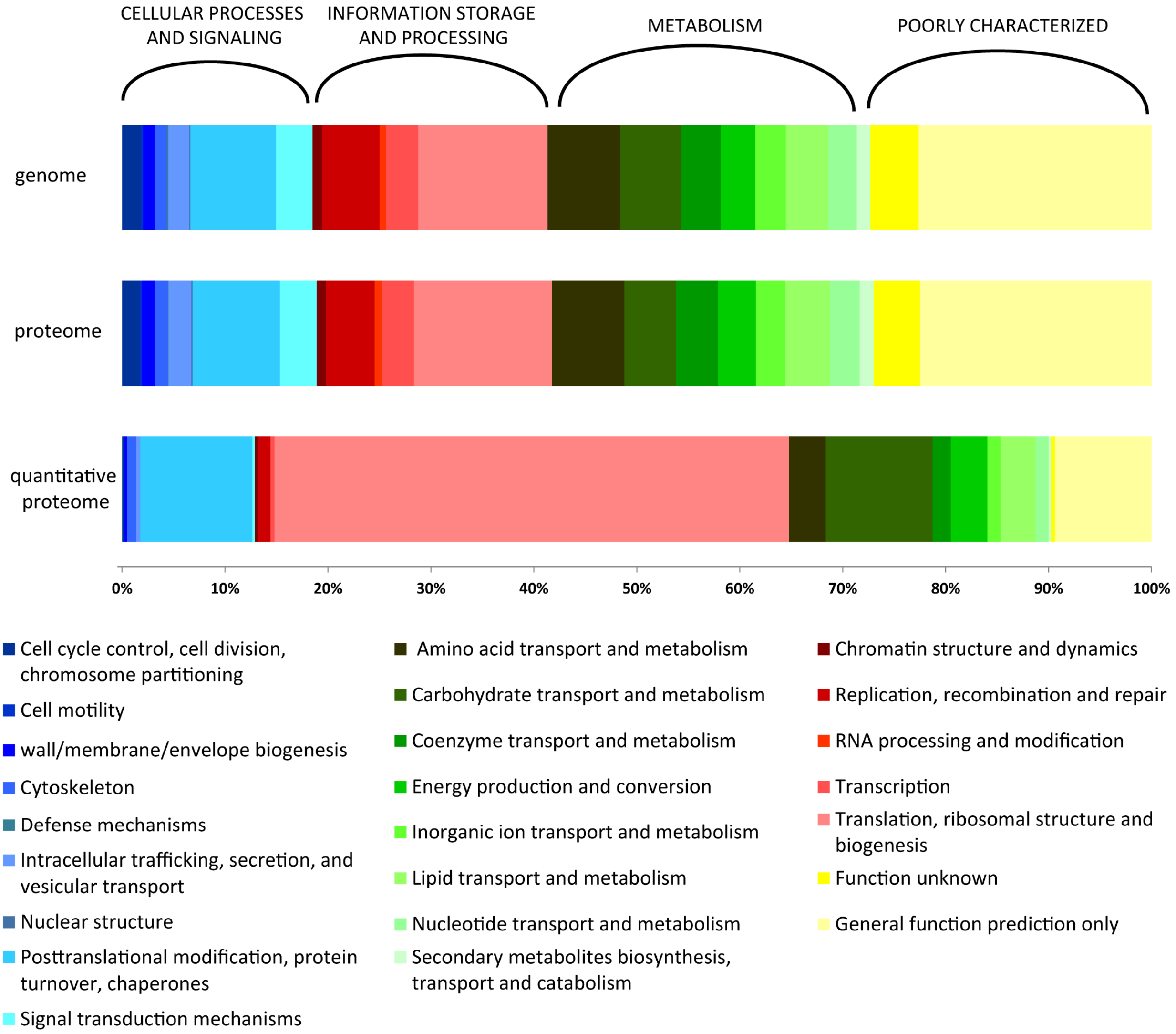


Figure5-15: Association of Clusters Orthologous Groups (COGs) classes composition found in the genome, proteome and quantitative proteome. The fraction of predicted genes and identified proteins are shown as percentages based on their biological category. The quantitative proteome displays the absolute protein levels in the appearance at eleven hours young mycelium. Functional COG categories are color coded in the legend. Genome n=2,047; Proteome/Quantitative proteome n=1,838.

To obtain a global picture of proteome rearrangements during cell growth, the protein profiles and abundances were determined for several growth states and in a system-wide manner. Clusters of Orthologous Groups (COG) describe functional characteristics of different microbe species (Tatusov et al., 2003). Figure5-15 shows the COG distribution of the *A. gossypii* genome, proteome (both by numbers), and the quantitative proteome (by mass). We analyzed the relationship between the genome, proteome, and the quantitative proteome. The genome and the proteome show huge similarities, which indicates a high identification rat of proteins across the whole genome. In contrast quantitative proteome shows huge differences to both, reviling the importance of expression regulation. The functional clusters provide a good representation of the actual allocation of protein classes in the cell, indicating that some proteins groups, which are low in number, are highly expressed and cover a large fraction of the quantitative proteome. Especially, proteins involved in posttranslational modification, carbohydrate transport, metabolism, and translation are highly expressed. In particular, translation contributes almost half of the protein mass in fast growing cells (11h time point). In contrast, low expressed proteins seem to be involved in the functional clusters of cell cycle control, chromatin structure and dynamics, and signal transduction. Moreover, proteins with unknown function are remarkably reduced in the quantitative proteome. Thus, these proteins are generally lower expressed and therefore probably less well-studied. These differences further enhance the importance of using quantitative proteomics to determine absolute protein abundances.

Table5-11: Cluster orthologous groups (COGs) dynamics during the three cell stage: spores, germ bubbles, and young mycelia.

Cluster orthologous groups (COG)	0h in [%]	5h in [%]	11h in [%]
Amino acid transport and metabolism	9.1	8.0	3.6
Carbohydrate transport and metabolism	9.7	11.9	10.3
Cell cycle control, cell division, chromosome partitioning	2.3	1.0	0.2
Cell motility	0.0	0.0	0.0
Cell wall/membrane/envelope biogenesis	1.1	0.6	0.2
Chromatin structure and dynamics	0.2	0.2	0.2
Coenzyme transport and metabolism	4.0	2.3	1.7
Cytoskeleton	1.5	2.1	0.9
Defense mechanisms	0.0	0.0	0.0
Energy production and conversion	7.9	11.3	3.6
Function unknown	1.5	0.9	0.4
General function prediction only	12.2	10.4	9.4
Inorganic ion transport and metabolism	3.3	2.2	1.2
Intracellular trafficking, secretion, and vesicular transport	1.8	0.8	0.4
Lipid transport and metabolism	4.9	4.6	3.5
Nuclear structure	0.2	0.1	0.0
Nucleotide transport and metabolism	2.8	1.4	1.3
Posttranslational modification, protein turnover, chaperones	10.8	13.3	10.9
Replication, recombination and repair	1.9	1.2	1.3
RNA processing and modification	0.1	0.1	0.0
Secondary metabolites biosynthesis, transport and catabolism	1.3	0.9	0.3
Signal transduction mechanisms	1.7	0.7	0.3
Transcription	0.6	0.4	0.4
Translation, ribosomal structure and biogenesis	21.2	25.7	49.9

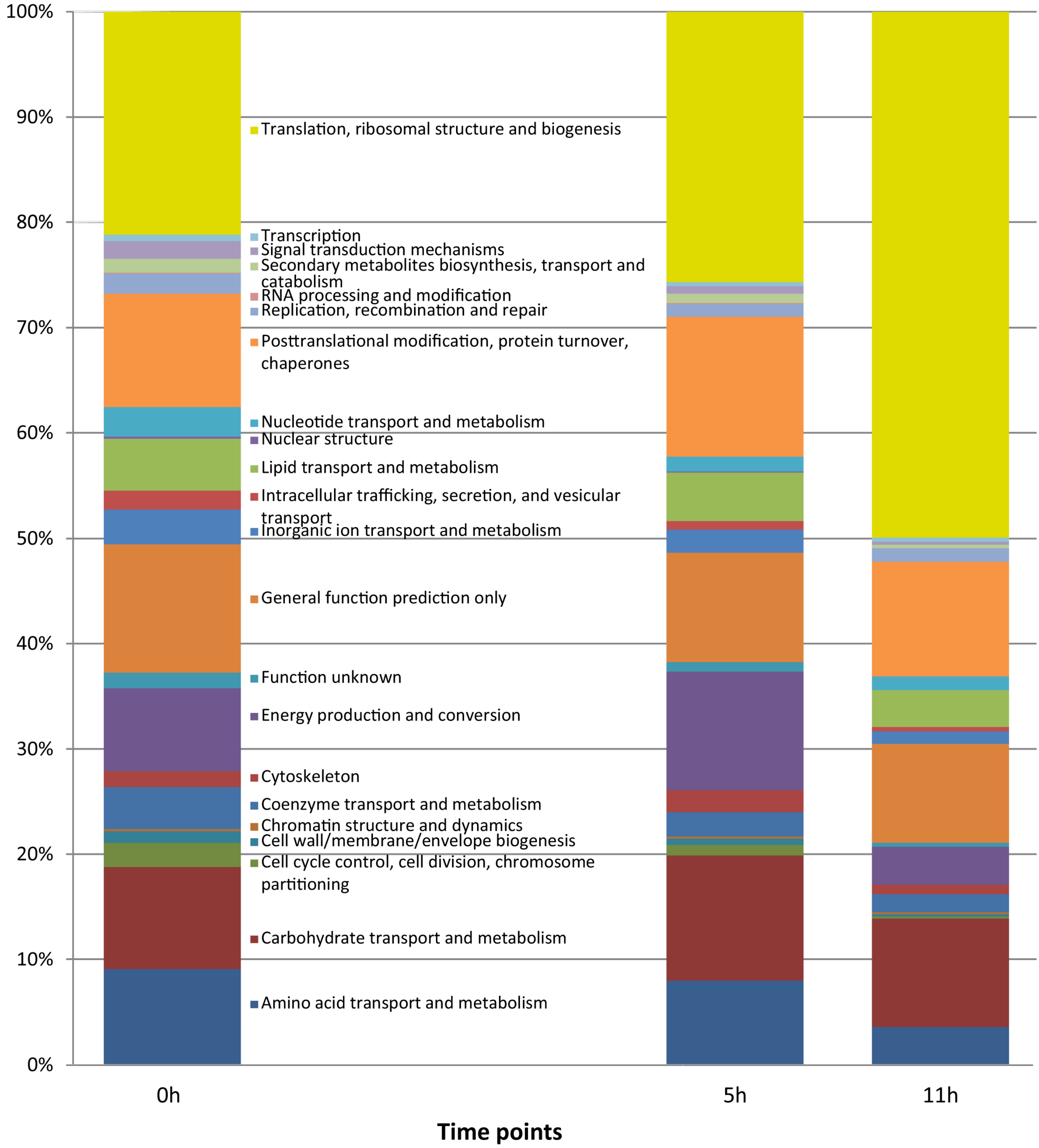


Figure5-16: Cluster orthologous groups (COGs) of predicted gene products are plotted in percent on the quantitative proteome levels. The fractions are defined by the functioning in specific biological processes (see legend).

In Figure5-16 the COGs protein fractions of different development stages are shown on the quantitative proteome levels in percent. The protein copy numbers of single proteins were calculated in masses *via* the molecular weight of the protein. The categories are shown in the insert from top to bottom (Figure5-16) and are fragmented by the function into specific biological categories. The COGs and the percentage of the protein mass of each group are listed in Table5-11.

The Translation, ribosomal structure and biogenesis cluster is the most abundant category in the total protein amount. In spores over 20% of the proteome is composed of proteins that belong to this category, which slowly increases to 25% after fife hours. The category including translation is highly induced after eleven hours. In total 50% of the quantitative proteome is represented by this class in young mycelium, which can explain the enormous biomass and protein concentration increase during the development. Another high represented COG is the group including posttranslational modification, protein turnover and chaperones. But in contrast to the previously described group this class shows a stable protein mass during the development of *A. gossypii*. The group of carbohydrate transport and metabolism is stable and prominent with 10% over the three time points. These processes are controlled by the nutrients source, which is needed to be transported in high amounts necessary for the accelerating growth speed. The total mass of proteins that are involved in the energy production and conversion shows first an increase in the quantitative protein amount from spores to germ bubbles and is in young mycelia reduced 3.6%. In full medium the carbon source glucose is present in high supply, which guaranteed energy maintenance and this functional category becomes less important to be present in such high percentages. The proteins organizing the amino acid metabolism are reduced after eleven hours from 9.1% to 3.6%, because of the growth in full media. Signal transduction mechanisms are reduced in the protein mass over time. Sporulation may be induced because of nutrients limitations and spores need to have a decent sensing of the environment to induce germination at right conditions, which can explain the higher amount of signal transduction proteins at zero hours. Signal transduction mechanisms can processes information of the surrounding environment, which is not needed in higher amounts for cells that are growing at optimized conditions. The protein quantifications of cytoskeleton components show an increase at 5 hours and decrease at 11 hours. Chromatin structure and dynamics are represented in low concentrations, but are stable (0.2%) during the development. Identified proteins with unknown functions have overall a low concentration and are reduced from 1.5% at zero hours to 0.4% at eleven hours. Proteins with unknown functions are expressed in low copy numbers in spores and have even less impacted in growing cells, which may be the reason why it is complicated to characterize these protein classes. Furthermore, proteins that regulates defense mechanisms and cell motility were not identified, which is in agreement with the immobility of *A. gossypii* cell and the lack of established defense mechanisms.

Clustering all quantified proteins in COG classes (Figure5-17) shows a general dynamics between spores, germ bubbles, and young mycelia. Figure5-16 shows the stacked bar diagram of this functional groups in including the no COG annotated quantified proteins (gray bar). We detected a general decrease in the protein mass of metabolic functions from 1/3 to 1/5 of the total proteome during the development. Proteins involved in information storage and processing are increased by a factor of 2.5 from spores to young mycelia of the quantified protein components, which is specifically caused by increase of ribosomal proteins (Figure5-16). Importantly, not all known *S. cerevisiae* proteins are annotated by COGs, which leads to a relative prominent amount of proteins in the empty category. In spores 33% of the quantitative proteome is not functional annotated by COGs. In young mycelia this amount decreases to 20%. Interestingly, COGs that are not annotated (Empty) decreases over time, which indicates that a high number of these proteins are not strongly expressed in growing cells, which leads to the suggestion that the majority of these proteins are not essential during growth. In total this does not change the characterization of the functional class behavior, but in some cases it can lead to changes of the protein cluster course.

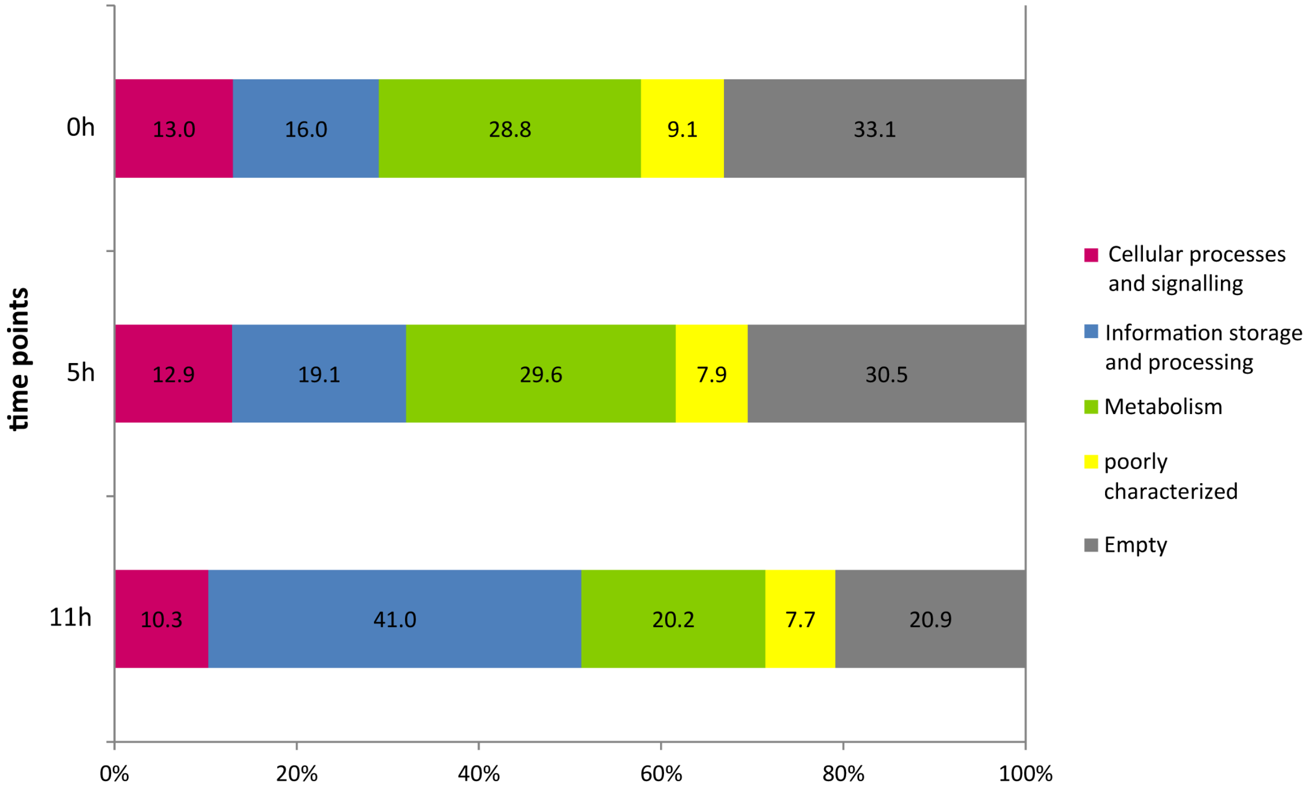


Figure5-17: Proteome composition changes in response to the germination process and hyphal growth. Cluster of orthologous group class dynamics over eleven hours of growth in liquid full medium. Functional classes are color coded as described in the insert.

Discussion

We confidently identify 4,217 proteins at a 1% false detection rate (FDR) based on the target-decoy search strategy (Elias and Gygi, 2007). Most proteins were confidently identified with five or more peptides (Figure5-18).

The Characterization of proteins during the development of *A. gossypii* via cluster analyses is an easy way to identify proteins that show same behaviors over certain time points. Followed by the systematic and integrative analysis with the DAVID Bioinformatics tool (Huang da et al., 2009) the identification of functional categories in this clusters is an informative way to show the regulation of different biological processes. The predictions of enriched protein groups in the clusters were identified in agreement with the biological understanding of development from spores to young mycelium. For example, ribosomal proteins are strongly induced during the germination and proteolysis genes are suggested to be repressed.

The Tables of the top 50 regulated proteins at different time points show extreme changes in the abundances. Most of the identified proteins in this list have low peptide coverage (fewer than three). This could explain why the values of these protein abundances are changing in such a high manner. High coverage of proteins followed over time is not crucial, because peptides are internally compared and have the same ionization frequency, in contrasted to comparing protein abundances of different protein species.

The Cluster of orthologous groups (COGs) based on the *S. cerevisiae* protein functions is in agreement with the prediction generated by the DAVID Bioinformatics tool and shows prominently similarities in the development. This finding we discovered as well by comparing the protein copy numbers and the protein abundances between different protein species (described in Chapter2 -4). We found that in young mycelia the majority of ribosomal proteins is highly enriched (Table4-1) where in spores other proteins are found to be more abundant (Table2-2).

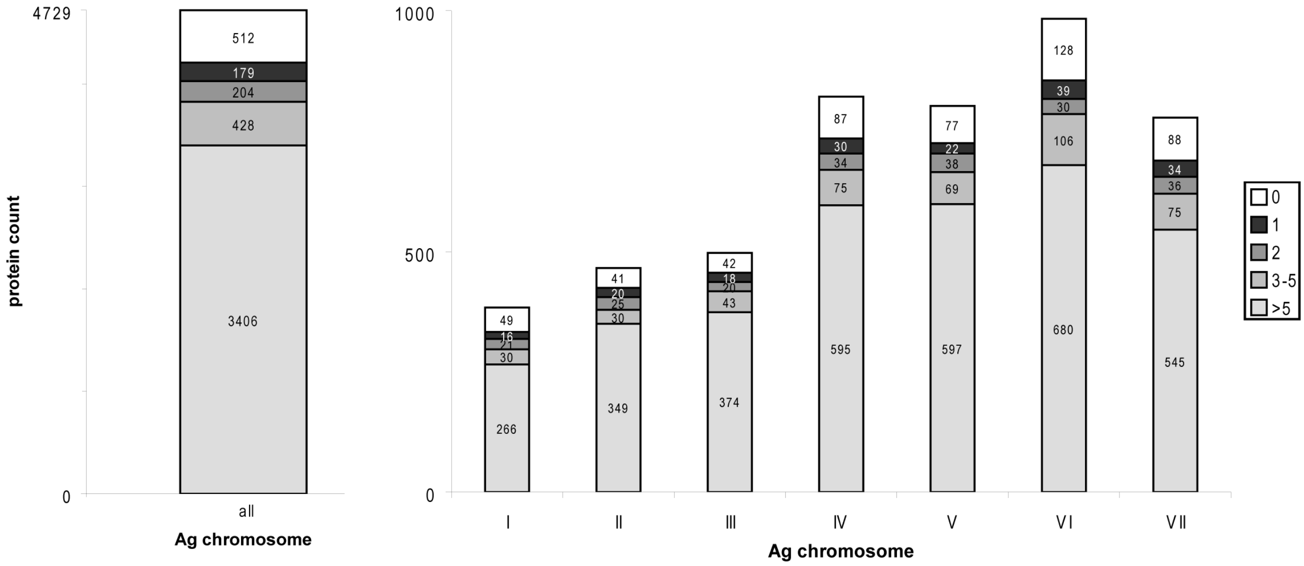


Figure5-18: All detected proteins based on the position in the genome during the proteome project of *A. gossypii*. White boxes represent genes, which products were not detected during this study. The numbers of detected peptides per protein are indicated.

The four septins

To demonstrate that the variations of quantified proteins during the development of *A. gossypii* are not caused by fluctuations of measurement we characterized the 4 septins (Table 5-12). The relative protein abundances and the protein copy numbers of the septins Cdc3, 10, 11, and 12 are in the same range and show the same abundances in the spore proteome (0h). By comparing the development over 11 hours, the septins show the same ratio and the same protein copy numbers per nucleus. The four septins (Hartwell, 1971) are known to be required for cytokinesis (Field and Kellogg, 1999). This uniform behavior of the septin protein complex suggests a hetero oligomer of four components, because these proteins are found to be in the same concentrations at different cell stages. Studies showed that septins are organized in multi-septin hetero-oligomers (Versele and Thorner, 2005) and in *S. cerevisiae* the core is proposed to assemble in the order of Cdc11–Cdc12–Cdc3–Cdc10–Cdc10–Cdc3–Cdc12–Cdc11 (Weirich et al., 2008). This is in agreement with our presented stoichiometric ratio based on the relative protein abundances. Our current knowledge of protein concentrations approves the septin organization and reflects the accuracy of the established MS data of the Ashbya gossypii proteome during the development from spores to young mycelia.

Table5-12: Development of the relative abundances and the protein copy number of the four septin proteins.

		relative abundance			absolute protein count			detected peptieds	possible peptides	protein size [aa]	size [MW]	Score
accession	common name	0h	5h	11h	0h	5h	11h					
AFR111C	Cdc3	1306	1731	398	25310	39578	6325	34	36	506	57358	1802
AAR001C	Cdc10	1260	1315	221	24392	30864	3547	33	23	328	37200	1697
AER445C	Cdc11	1426	931	296	27718	22587	4726	20	26	411	46791	1191
AER238C	Cdc12	2230	2049	335	43974	46100	5349	40	29	390	44751	2101

Materials and methods

Materials and methods

Strains and growth conditions

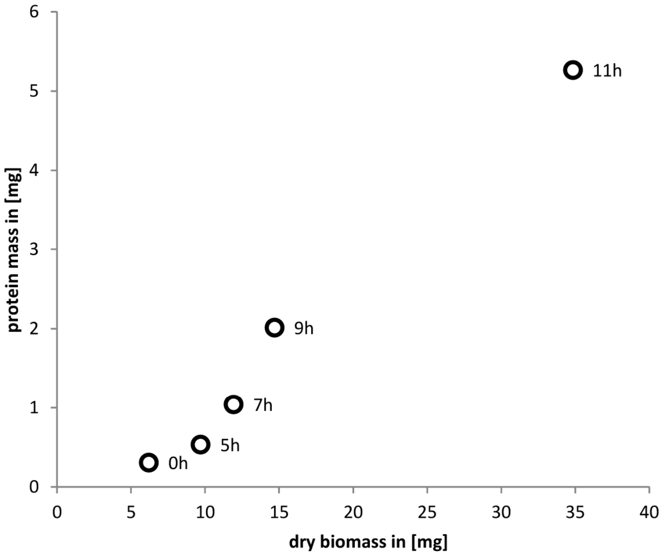
For this study we used the laboratory strain Agleu2Δthr4Δ, a derivative of the *A. gossypii* wild-type strain ATCC 10895 (Ashby and Nowell, 1926), which carries deletion of the AgLEU2 and AgTHR4 genes (Altmann-Johl and Philippsen, 1996). *A. gossypii* was grown on full medium (AFM) containing 10 g/l Bacto Peptone (Pancreatic Digest of Casein, Difco), 10 g/l Yeast Extract (Micro Granulated, Formedium, Norwich, England) 1 g/l Myo-inositol (Merck), and 2% D(+)-Glucose-Monohydrate (Merck). Glucose was autoclaved separately and added directly before usage to the medium. For spore production, *A. gossypii* was grown on 35x10 mm Petri dish (Falcon 1000, Becton Dickinson Labware, New Jersey, USA) by adding 15 g/l agar (Formedium, Norwich, England) to solidify the medium. Cultures were incubated at 30°C in a temperature controlled environment. Liquid cultures were specially incubated in baffled flasks (Duran, Schott, Germany) on a rotary shaker (INFORS® AG) at 150 rpm, to guarantee oxygen supply.

Spore preparation

Spores were isolated without enzymatic digestion by exploiting their high hydrophobicity to eliminate vegetative mycelial contamination. We used coated glass tubes with Sigmacote® (Sigma-Aldrich CAT# SL2) which were incubated overnight and autoclaved to fix the cote to the glass. The method was invented and described by Sophie Lemire-Brachat (Brachat, 2003). Harvested mycelium containing the spores was incubated in the coated glass tube with dH2O for several hours and after washing three times with dH2O, the spores were eluted with 0.1% (v/v) Triton X-100 (Triton® X-100 BioChemica, AppliChem, CAT# A1388,1000) by vortexing under strong shaking conditions. Spores were transferred from a 15 ml Falcon tubes into a 1.5 ml Eppendorf tubes by centrifugation at 3000 rpm for 5 minutes at room temperature, washed and incubated overnight at 4°C with 0.03% (w/v) RapiGest SF Surfactant (Waters, CAT# 186002123). All spore preparations were controlled by light microscopy for purity and only fresh spores were used as an inoculums for liquid cultures and the sampling of life cycle stages. For more information and detailed description see chapter2 part A and figure2-2/2-3.

Sampling of life cycle stages

We obtained different life cycle stages of the fungus *A. gossypii* by inoculation of fresh spores in liquid full medium (AFM) and assayed them after: 0h, 2h, 5h, 7h, 9h, 11h, and 18h. For the sample collection at 0h time point we used 169 mg of fresh spores originated from 40 six days old *A. gossypii* colonies on AFM plates and the use of 60 coated glass tubes to produce 5.4 mg protein extract for LC-MS analysis (for more details see Chapter2 partA). For the other time points, we followed the established diagram of the protein concentration based on the biomass and cell population stage in Figure5-3 and Figure6-1 to estimate spore levels. We prepared a uniform spores solution in AFM, generated equal aliquots in 250 ml baffled flasks and grow them for 0, 5, 7, 9, 11, and 18 hours. Time point samples were dried using a CentriVap Vacuum Concentrator System (Labconco). The weighted samples were then lysed in lysis buffer (8 M urea, 0.1% RapiGest, 0.1M ammoniumbicarbonate). Protein concentrations were determined by Bradford assay (Bradford, 1976). Figure6-1 shows the increase of the dry biomass and protein mass by an inoculation of 6.192 mg spores into AFM. To control the different protein yields, we adjusted the spore amount that was inoculated for the different time points accordingly. To ensure optimized nutrients and oxygen supply for the growth we used 250 ml baffled flasks filled with 50 ml AFM for the 0h, 2h, and 5h time points and for the 7h, 9h, 11h, and the 18h time point we used 200 ml AFM in 500 ml baffled flasks. Additionally, to growth higher amounts under optimal conditions, we used multiple baffled flasks per time point. 500 µl aliquots were removed and fixed with paraformaldehyde and stored for later DAPI staining and inspection. Sample collection was done via centrifugation at 4000 rpm and room temperature in 50 ml falcon tubes and stored in 2 m micro tubes (Sarstedt CAT# 72.693.005) at -80°C after fast freezing in liquated nitrogen.



time point	biomass in mg	protein mass in mg
0h	6.192	0.310
5h	9.692	0.535
7h	11.917	1.043
9h	14.692	2.016
11h	34.842	5.267

Figure6-1: Correlation of needed biomass of specific cell stages to gain protein amounts. All time points (0, 5, 7, 9, and 11 hours) were inoculated with the same amount of spores (6,2mg) to define the mg of spores is required to reach a cell stage biomass and a extracted protein amount.

Cell fixation

500 µl of the time point sample were mixed with 4°C cold 500 µl 4% paraformaldehyde (PFA) solution and incubated for 20 min at room temperature on ice. Fixed cells were then washed three times with 1xPBS (4 mM K_2HPO_4; 1 mM KH_2PO_4; 15 mM NaCl) and diluted with 0.03% Triton X-100. The washing was performed via centrifugation with a „Biofuge Pico“ centrifuge (Heraeus) at 500 rpm for five minutes. Fixed samples were stored at 4°C. The 4% PFA was prepared by dissolving 4 g PFA with 50 ml sterile H_2O and 1 ml 1M NaOH at 65°C. Afterwards, 10 ml 10x PBS (0.04M K_2HPO_4; 0.01 M KH_2PO_4; 0.15 M NaCl) were added and after cooling down to room temperature the pH was adjusted to 7.4. Finally, the volume of the solution was adjusted to 100 ml with sterile H_2O.

DAPI staining

To stain the nuclei of *A. gossypii* we used the blue fluorescent dye 4,6-Diamidino-2-phenylindol (DAPI) (CarlRoth, CAT# 6335.2). To a total amount of 200 µl fixed cells in PBS, 0.03% Triton X-100 we added 1 µl of a 1 mg/ml DAPI ethanol solution and incubated on ice in the dark for 10-20 min. Cells were washed 1x with PBS and 3-5 µl of stained cells were placed on a microscope slide and sealed with a cover slip.

Microscopy and fluorescence microscopy & image processing

For microscopy we used an Axioplan2 Imaging-System equipped with an automated Epi-fluorescent microscope (Carl Zeiss AG, Feldbach, Switzerland), which is controlled via MetaMorph 6.2r5 software (Molecular Devices Corp., Downingtown, PA). For the setup, objectives Plan-Apochromat 100x/1.40 NA Oil DIC and Plan-Apochromat 63x/1.40 NA Oil DIC and appropriate filters (Zeiss and Chroma Technology, Brattleboro, VT) were used. Images were acquired using a cooled charge-coupled device camera CoolSNAP HQ (Photometrics, Tucson, AZ) at room temperature. The microscope was setup according to (Rodicio et al., 2006) and (Knechtle et al., 2003). Z-distance of two planes was set between 0.1 and 0.5 µm. Image processing was performed with MetaMorph 6.2r5 software. Z-stacks were compressed by maximum projection with Stack Arithmetic and converted from 16-bit to 8-bit and stored as 8-bit gray scale and RGB TIFF files. Following processing and measurements were although processed with ImageJ (Wayne Rasband, National Institute of Health).

Bradford & BCA – assay

We used Bradford & BCA assay to determine the protein concentration of the samples. For the Bradford measurements (Bradford, 1976), we used the protein assay from BIO-RAD (München) and attend the orders of the producer. The measurements were performed with a photometer Ultraspec 4000 (Pharmacia Biotech) at a wave length of 595 nm. The concentration definition was done via a regression line by the use of bovine serum albumin (BSA). The BCA assay of Pierce® BCA Protein Assay Kit (CAT# LE145476) is based on bicinchoninic acid (BCA) for color detection and quantification of total protein. The method was performed according to the supplier instructions.

SDS-Polyacrylamid Gelelektrophorese (SDS-PAGE)

For sample quality control and digestion completeness, we use discontinuous sodium dodecyl sulfate-Polyacrylamide-Gel-Electrophorese (SDS-PAGE) to separate proteins based on their molecular weight (MW) (Laemmli, 1970). We used the 2D Gel electrophoresis system (Amersham Bioscience Hoefer SE 400 Series) with a 12% resolving gel (10.42 ml dH_2O; 6.5 ml 1.5M Tris/HCl, pH 8.8; 8.66 ml 30%Acryl Amid; 260 µl 10%SDS; 120 µl 10%APS; 40 µl TEMED) and a 5% stacking gel (5.3 ml dH_2O; 2.5 ml 0.5M Tris/HCl, pH 6.8; 2 ml 30%Acryl Amid; 100 µl 10%SDS; 60 µl 10% ammonium persulfate (APS); 20µl TEMED). We proceeded with the gel system and cast gel according to the instructors manual and separated the proteins in running buffer (1.2%Tris base, 5.76% glycerin, 1% SDS, pH 8.3). 20 µl of each sample were 1:1 mix with fresh sample buffer (0.355 ml dH_2O; 0.125 ml 0.5M Tris-HCl, pH 6.8; 0.25 glycerol, 0.2 ml 10%(w/v) SDS; 20 µl 0.5%(w/v) Bromophenol Blue; 50 µl β-mercaptoethanol) and heated for 5 min at 95°C. We used the prestained protein ladder mix from Invitrogen (MultiMark) and the BioRad protein ladder mix (Dual Color). The experimental parameters were applied according to the descriptions by (Sambrook and Russel, 2001). The proteins were aligned in stacking gel at a current of 15 mA and separated in the Resolving gel at 25 mA. To visualize the proteins we stained the gel with Coomassie Brilliant Blue (10% (v/v) acetic acid; 40% (v/v) EtOH; 50% (v/v) ddH_2O; 1-2 g Brilliant Blue) for several hours, washed them multiple times in washing buffer (10% (v/v) acetic acid; 30% (v/v) EtOH; 60% (v/v) ddH_2O) and scanned the stained gel.

Sample preparation

Protein extraction from whole cells

The collected samples were split into aliquots of about 200 mg mycelia or spores, transferred in 2 ml bead beater tubes (Sarstedt CAT# 72.693.005) and dissolved in 500 µl ice cold lyses buffer (8M urea; 0.1%(w/v) RapiGest, 0.1M ammonium bicarbonate) containing 500 µl of chilled 0.5 mm Zirconia/Silica beads (BioSpec Products Inc. CAT# 11079105z). The first steps of the cell lysis was done via the FastPrep® Cell Disrupter (FP220A Instrument, MP Biomedicals Europe) by applying full speed (6.5) and six cycles of 30 sec each, with sample placement on ice for 2 minutes between each cycle. The extract was collected by puncturing the bottom of the tube with a needle and insert the bead beater tube into an 2 ml Eppendorf tube, followed by quickly accelerate centrifuge at 3000 rpm at 4 °C by placing the stacked tubes in a 50 ml falcon tube. Afterwards the Zirconia/Silica beads were washed with additional 200 ml lysis buffer. The protein extracts were then sonicated (Hielscher Ultra sound Technology) to destroy remaining parts of the lipid bilayer and dissolve membrane proteins in the samples. Three to six cycles depending on the density with the highest ultrasound energy were applied to each sample. The lysate was cleared by 5 minute centrifugation at 14000 rpm at 4°C and the supernatant transferred to a new 2 ml tube. The proteins in the extract were then precipitated to get rid of Triton X-100 that would interfere with MS analysis.

TCA protein precipitation

For the protein precipitation we added 1 volume of 100% (w/v) Trichloroacetic acid (TCA) to 4 volumes of protein sample and incubated for 10 min at 4°C. The sample was spun down in a micro centrifuge at 14000 rpm for 5 min. Supernatant was removed and the intact protein pellet was washed two times with 200 µl cold acetone. The protein pellet was dried at 95°C on heat block for 5 to 10 min and dissolve in 350 µl lyses buffer by shaking on an Eppendorf Thermomixer combined with Sonication. The final protein concentration was measured by BCA assay.

Reduction and alkylation of disulfide bonds

For the reduction and alkylation of disulfide bonds, we adjusted the protein concentration to 2.5-10 mg/ml and corrected the pH to 8.5 using 0.1M HCl / NaOH. We added 1 µl 0,2M tris(2-carboxyethyl)phosphine (TCEP) (57.23 mg/ml TCEP dissolved in 0.1M TRIS pH=8.5) per 40 µl protein extract and kept the mixture at 37°C for 1 h at 500 rpm on a Eppendorf Thermomixer comfort. After the mixture was cooled down, 1 µl per 40 µl sample 0.4M freshly prepared iodoacetamide solution (74 mg/ml dissolved in water) were added and kept at 25°C for 30 min at 500 rpm in the dark. Next we pipetted 1 µl per 40 µl sample 0.5M N-acetyl-cysteine (81.6 mg/ml dissolved in 0.1M TRIS pH=8.5) to the mixture and after vortexing we left the sample at room temperature and 500 rpm for another 10 min.

Proteolysis

For protein digestion we mixed the alkylated samples 1:4 with 0.1M ABC buffer to get a final urea concentration below 2M. The pH was adjusted to 8-8.5 using 0.1M NaOH / HCl and trypsin (Sequencing Grade Modified Trypsin, Promega CAT# V5111) was added to a final enzyme / protein ration of 1:50. The overnight digestion at 37°C was controlled by SDS PAGE to ensure complete proteolysis and, if incomplete, additionally trypsin was added and incubated another for several hours. Post digestion, 5% TFA was added to a final sample concentration of 0.5% and HCl was added to a final concentration of 50mM. The adjustment was done to gain a pH lower than 2 and to cleave RapiGest. Samples were then incubated at 37°C for 45 min at 500 rpm and centrifuged at 14000 rpm for 10 min at 4°C. The supernatant was transferred into a new 2 ml tube. AQUA peptides were spiked into the sample at this stage where applicable using the described concentrations and amount discussed in Chapter2 part C. In brief, AQUA peptides were grouped into two abundance classes and spiked into *A. gossypii* peptide mixtures directly after digestion, at a final concentration of 2 pmol or 20 pmol per 100 µg of protein extract.

Solidphase extraction and speed-vac concentration

For peptide purification and preparation for the MS we used C18 columns (Sep-Pak from Waters and Micro Spin Columns C18 Silica from The Nest Group, Inc). The peptide cleaning with the C18 reversed-phase spin columns was performed according to the manufacturer‘s instructions. The eluted peptides were dried under vacuum (CentriVap Vacuum Concentrator System, Labconco) and dissolved in LC-MS/MS buffer A (98% H_2O, 2% acetonitrile, 0.15% formic acid) using 2 minutes ultra-sonication and shaking at 1400 rpm at 37°C for 5 min. Samples were stored at -20°C until further use.

OffGel Electrophoresis (OGE)

The peptides were separated on pH 3–10 ImmobilineTM Dry Strips (GE Healthcare) with a 3100 OFFGEL fractionator (Agilent Technologies). The sample and stock solutions were prepared according to the producer instructions. For the peptide library we used 24 cm IPG strips and for the high resolution time point we used 13 cm IPG strips both having a pH range of 3-10. The 24/12 fractions collected from the 24/13 cm IPG strip separation were desalted using Micro Spin Columns (Nest Group), respectively, before LC-MS analysis. For more details see Chapter 1.

Mass spectrometry

The set-up of the MS system was done as described in (Schmidt et al., 2008). Each fraction was analyzed in a combined shotgun and directed (inclusion list) MS method to specifically directed LC-MS/MS analysis to the selected spiked in heavy reference and endogenous peptides according to Schmidt et al., 2008. Samples for absolute protein quantification were analyzed using nLC/Orbitrap-Velos (ThermoScientific, Bremen, Germany) LC-MS system with the following modified parameters: peptides were separated using a linear gradient from 92% solvent A (98% water, 2% acetonitrile, 0.15% formic acid) and 8% solvent B (98% acetonitrile, 2% water, 0.15% formic acid) to 40% solvent B over 120 min. Each survey scan acquired in the Orbitrap at 60 000 FWHM was followed by MS/MS scans of the 10 most intense and additional 3 precursor ions matching inclusion list masses in the linear ion trap (Beck et al., 2011).

Protein identification and quantification

The acquired raw-files were imported into the Progenesis LC-MS software (v3.0, Nonlinear Dynamics Limited), which was used to extract peptide precursor ion intensities across all samples applying the default parameters. The generated mgf-files were searched using MASCOT against a decoy database (consisting of forward and reverse protein sequences) of the predicted proteome from 4,748 *A. gossypii* proteins. The database consists of 4,748 *A. gossypii* proteins as well as known contaminants such as porcine trypsin, human keratins and high abundant bovine serum proteins (Non-Redundant Protein Database, National Cancer Institute Advanced Biomedical Computing Center, 2004). The search criteria were set as follows: full tryptic specificity was required (cleavage after lysine or arginine residues, unless followed by proline); 2 missed cleavages were allowed; carbamidomethylation (C) was set as fixed modification; oxidation (M), acetylation and mono-di-trimethlyation (K) were applied as variable modifications; mass tolerance of 10 ppm (precursor) and 0.6Da (fragments). The database search results were filtered using the ion score to set the false discovery rate (FDR) to 1% on the peptide and protein level, respectively, based on the number of reverse protein sequence hits in the datasets. The quantitative data obtained were further normalized and statistically analyzed according to Brusniak et al., 2008.

Absolute protein quantification

Absolute quantification was carried out according to Schmidt et al., 2011. In brief, the raw files were converted to mzXML file format, the database searched using XTandem/PeptideProphet followed by isotope ratio calculation by the Xpress software tool. All software tools are part of the trans-proteomics pipeline (Deutsch et al., 2010). The median peptide ratios were used to determine the endogenous protein levels in the individual samples. Based on the number of disrupted cells counted in triplicates for each sample before and after cell lysis, absolute abundances for the selected proteins (in copies/cell) could be calculated across all samples. These were aligned with the summed protein or top3 peptide intensities as provided by the Progenesis LC-MS software (v3.0, Nonlinear Dynamics Limited) divided by the number of expected tryptic peptides for the summed intensity values. The thus generated models were applied to estimate absolute protein levels for all quantified proteins in the individual samples.

Appendix

Appendix 1:

Protein abundances of the three time points: 0h, 5h, and 11h (Table2-2)

Protein abundances of the three time points (0 h, 5 h, and 11 h) determined with LC-MS/MS and high resolution OGE fractionation were combined in Table2-2. The complete Table2-2 is attached in an electronic *.xlsx file. Results are discussed in detail in Chapter2-4. Developmental changes of protein abundances between the time points were discussed in Chapter5. Table2-2 is ordered by the protein abundances starting with the most abundant protein at 0h using the All-Method. The printed version below is an example of the organization of the complete Table2-2 and is sorted by the *A. gossypii* accession entries (column1). The common names are based on the homologs in *S. cerevisiae* (column2). Furthermore, protein abundances were determined using the All-Method (*1) (column3-5) and the Top3-Method (*2) (column6-8). We included the detected and possible peptides per protein (column9+10), the amino acid length (column11), and the molecular weight (column12). Additionally, we determined the protein scores based on the sum of the MASCOT peptide scores (column13).

Table2-2

		relative abundance (All *^{1})			relative abundance (Top3 *^{2})			detected	possible	protein	protein	
accession	common name	0h	5h	11h	0h	5h	11h	peptides	peptides	size [aa]	size [MW]	Score
AAL001W	Dom34	27.5	48.9	19.7	214	380	145	13	31	385	43576	551
AAL002W	Rer1	419	229	48.7	1635	847	182	7	12	183	22074	288
AAL003W	Pgs1	18.6	13.8	3.65	153	105	34.5	12	39	508	57897	525
AAL004W	Ldb16	0.799	0.0082	0	10.8	0.110	0	2	27	311	35020	55
AAL005W	Vma9	184	29.5	53.5	368	59.0	107	2	4	72	8287	52
AAL006C	Stp22	29.9	19.4	1.18	245	161	8.62	6	26	445	50315	308
AAL007C	Cbf5	239	500	1220	1440	2603	5687	37	29	485	54782	1864
AAL008W	Sis1	385	264	643	1413	767	1986	31	19	349	37744	1521
AAL009C	Lst8	89.7	68.3	6.07	576	473	39.5	8	22	303	33893	406
AAL010W	Rfx1	0.306	0.961	8.82	3.26	10.2	88.4	6	32	834	92365	207
AAL011C	Rlp7	26.5	113	385	125	390	1331	11	16	293	32527	635
AAL013W	Nfs1	1127	569	221	9769	4196	1536	27	29	490	54384	1468
AAL014C	Pet8	2.52	15.5	8.46	12.2	69.2	42.0	7	15	271	29832	274
AAL015W	Dcc1	1.71	2.45	6.64	14.1	18.2	50.4	7	26	363	40331	265
AAL016C	Bud3	0.956	0.312	3.60	29.3	9.08	72.4	9	95	1478	165582	398
AAL017W	YCL011C-A	5.96	2.60	9.39	65.5	28.6	103	1	11	104	11722	60
AAL018W	Hrb1, Gbp2	15.8	17.7	42.7	119	107	289	7	23	337	37647	335
AAL019W	Sgf29	0.339	2.69	8.63	1.69	10.7	34.6	5	15	236	26239	176
AAL020C	Mrp7	64.4	117	130	325	471	508	22	26	359	40978	1207
AAL021W	Ilv6	365	417	248	1617	2057	859	13	18	300	32604	868
AAL022W	Idp2, Idp3	3286	2704	573	20726	14597	3481	42	34	415	46536	2180
AAL023W	YNL010W	123	99.1	66.9	577	377	225	18	18	295	33102	856
AAL024C	YNL011C	85.5	38.1	5.83	635	283	43.3	6	23	444	49195	358
AAL025W	Ssp120	808	229	344	3165	767	1066	15	15	228	26337	817
AAL028W	Yef3, Hef3	2136	3610	7757	13978	19029	48173	112	63	1044	115757	6495
AAL029W	Rck2, Rck1	21.4	60.4	24.8	180	419	171	16	32	534	59489	747
AAL033W	Trm5	39.1	28.4	59.7	359	187	454	16	35	489	54949	847
AAL034C	Rrp4	124	174	197	731	870	1123	15	20	351	38784	774
AAL035W	Oms1	20.6	57.0	49.5	181	497	513	10	32	452	52165	446
AAL036W	Dys1	1217	1864	775	4999	8993	3020	29	21	382	41939	1700

Appendix 2:

Protein abundances of the time course from zero to nine hours (Table6-1)

Table6-1

	Normalized abundance															Peptide count	Confidence score	Max fold change
Accession	0h-1	0h-2	0h-3	2h-1	2h-2	2h-3	5h-1	5h-2	5h-3	7h-1	7h-2	7h-3	9h-1	9h-2	9h-3			
AAL002W	5.36E+05	1.72E+06	6.49E+05	2.65E+06	1.66E+06	7.77E+05	9.46E+05	3.13E+05	9.75E+05	1.54E+06	9.13E+05	9.17E+05	9.11E+05	1.14E+06	1.16E+05	2	5.0	2.3
AAL005W	3.48E+05	3.94E+05	3.16E+05	4.96E+05	2.14E+05	3.32E+05	3.12E+05	4.18E+05	2.18E+05	3.16E+05	1.02E+05	1.22E+05	7.56E+04	4.26E+05	2.39E+05	1	2.5	2.0
AAL007C	7.15E+06	6.72E+06	8.20E+06	3.92E+06	8.08E+06	1.03E+07	9.83E+06	7.90E+06	1.40E+07	5.84E+06	3.43E+07	2.38E+07	3.71E+07	2.27E+07	1.87E+07	20	98.8	3.6
AAL008W	7.29E+06	7.15E+06	6.18E+06	2.68E+06	7.37E+06	5.90E+06	4.06E+06	1.37E+07	9.10E+06	5.56E+06	9.01E+06	8.28E+06	9.76E+06	7.57E+06	7.58E+06	13	54.0	1.7
AAL009C	8.43E+04	2.11E+05	1.07E+05	1.75E+05	1.21E+05	2.29E+05	1.29E+05	5.40E+04	2.72E+05	3.47E+05	3.70E+05	2.61E+05	6.40E+04	7.18E+05	3.61E+05	1	2.8	2.8
AAL011C	1.37E+07	7.44E+05	1.01E+06	8.41E+05	8.54E+05	1.12E+06	1.39E+06	1.13E+06	1.32E+06	2.76E+06	3.86E+06	1.97E+06	3.34E+06	1.17E+07	6.26E+06	6	23.1	7.6
AAL013W	1.05E+06	1.99E+06	1.95E+06	3.64E+06	7.73E+05	4.76E+06	1.34E+06	8.77E+05	5.34E+06	3.38E+06	4.60E+06	3.86E+06	4.39E+06	1.77E+06	4.15E+06	7	27.5	2.4
AAL014C	1.07E+05	2.63E+04	1.34E+04	1.88E+04	8.09E+04	3.45E+03	1.13E+04	5.55E+04	2.96E+04	9.10E+03	5.65E+04	5.31E+03	8.85E+04	6.57E+03	1.80E+03	1	3.3	2.1
AAL020C	1.68E+06	1.07E+06	1.80E+06	1.18E+06	8.75E+05	2.29E+06	9.98E+05	8.58E+05	2.27E+06	8.94E+05	1.98E+06	2.47E+06	3.08E+06	1.82E+06	1.97E+06	9	23.5	1.7
AAL021W	8.29E+05	8.47E+06	4.22E+06	1.17E+06	2.19E+06	6.32E+06	2.20E+06	2.34E+06	7.59E+06	1.96E+06	3.46E+06	6.26E+06	5.65E+06	2.03E+06	1.55E+07	8	34.8	2.4
AAL022W	3.22E+07	3.36E+07	7.39E+07	3.06E+07	3.17E+07	8.95E+07	3.81E+07	2.47E+07	7.49E+07	1.31E+07	4.59E+07	6.29E+07	5.40E+07	1.97E+07	2.19E+07	25	96.6	1.6
AAL023W	1.18E+05	3.78E+05	3.20E+05	1.91E+05	2.29E+05	9.36E+05	1.11E+06	3.03E+05	6.83E+05	5.21E+05	1.01E+06	7.76E+05	3.99E+05	1.25E+06	2.22E+05	1	6.0	2.8
AAL028W	1.04E+08	1.48E+08	5.68E+08	7.55E+07	2.34E+08	4.02E+08	1.78E+08	3.42E+08	5.65E+08	1.41E+08	6.94E+08	7.90E+08	1.01E+09	4.32E+08	3.52E+08	58	345.1	2.5
AAL029W	3.09E+03	3.94E+04	6.97E+04	2.03E+04	7.46E+04	4.12E+04	7.38E+04	7.47E+04	4.28E+04	4.11E+04	9.55E+04	3.51E+04	5.11E+04	4.11E+04	1.37E+04	1	2.5	1.8
AAL033W	5.97E+04	1.02E+05	2.75E+05	1.24E+05	1.17E+05	3.13E+05	1.74E+05	1.52E+05	4.03E+05	6.18E+04	2.28E+05	5.07E+05	5.64E+05	5.94E+04	8.50E+04	2	9.3	1.8
AAL035W	4.32E+05	7.59E+05	5.89E+05	7.06E+05	2.74E+05	6.79E+05	8.42E+05	8.42E+05	7.94E+05	9.13E+05	8.29E+05	8.46E+05	8.71E+05	1.79E+06	2.31E+05	1	4.4	1.7
AAL036W	3.12E+07	3.50E+07	3.86E+07	7.35E+06	4.68E+07	3.80E+07	1.75E+07	2.50E+07	3.24E+07	2.11E+07	1.76E+07	3.05E+07	2.45E+07	2.20E+07	2.12E+07	17	78.7	1.5
AAL040W	1.11E+04	9.13E+03	6.71E+03	2.03E+04	2.49E+04	6.78E+03	1.60E+04	4.45E+03	1.69E+04	9.32E+03	7.15E+04	4.13E+04	8.56E+04	1.56E+05	3.76E+04	3	8.5	10.4
AAL041C	1.61E+04	4.99E+04	1.24E+05	3.07E+04	7.01E+04	1.03E+05	9.08E+04	1.46E+05	2.68E+05	1.15E+05	6.94E+05	3.33E+05	5.79E+05	3.27E+05	1.87E+05	2	5.3	6.0
AAL043C	4.52E+06	6.03E+06	1.26E+07	4.54E+06	4.74E+06	1.74E+07	5.40E+06	5.83E+06	2.22E+07	4.71E+06	2.54E+07	2.83E+07	4.69E+07	1.28E+07	1.73E+07	19	81.2	3.3
AAL049C	5.84E+03	2.27E+04	3.40E+03	1.34E+04	7.54E+03	3.31E+04	1.55E+04	2.22E+04	3.86E+04	2.81E+04	4.23E+04	4.32E+04	4.92E+04	4.52E+04	7.87E+04	1	5.0	5.4
AAL053C	2.30E+04	3.29E+04	8.10E+03	9.45E+03	1.86E+04	1.78E+04	5.17E+03	1.69E+04	2.15E+04	1.39E+04	1.22E+04	5.94E+03	1.48E+04	8.08E+05	9.49E+03	1	1.5	26.0
AAL054W	2.70E+04	9.18E+04	8.53E+03	7.49E+04	7.05E+04	7.73E+03	2.53E+04	1.90E+04	2.33E+04	8.62E+04	5.14E+04	1.40E+04	2.53E+04	6.10E+04	6.58E+03	1	3.0	2.3
AAL056C	1.49E+06	1.21E+06	9.23E+05	4.50E+05	1.07E+06	2.11E+06	1.07E+06	1.86E+06	1.92E+06	8.33E+05	1.41E+06	2.17E+06	1.82E+06	1.68E+06	5.28E+05	3	10.2	1.3
AAL061C	2.35E+06	3.12E+06	8.30E+06	3.30E+06	5.23E+06	8.95E+06	6.94E+06	5.82E+06	6.63E+06	3.77E+06	5.38E+06	6.18E+06	5.99E+06	6.47E+06	2.15E+06	7	30.1	1.4

To establish the best suited time points for our developmental study, we first checked short time intervals (called time course) with low resolution LC-MS/MS runs. This time course for the first steps in *A. gossypii* development, consisted of samples taken after 0 h, 2 h, 5 h, 7 h, and 9 h. For these LC-MS/MS analyses we used three biological replicas for each time point. The results of all data points are combined in Table6-1 sorted by the *A. gossypii* accession entries. A subset of Table6-1 is shown in this Appendix. The complete Table6-1 is attached as an electronic *.xlsx file. Protein abundances were calculated by sum up all peptides which are corresponding to one protein resulting in the normalized abundance. In total, we determined about 2,000 proteins with a false discovery rate (FDR) of 1%. Additionally, we listed the peptide numbers, the confidence scores, and the maximal fold changes between the five time points.

Appendix 3:

Protein library of *A. gossypii* mycelia (Table6-2)

In the proteome study of *A. gossypii* we first generated a protein library, to see how many proteins can be detected matching the predicted proteome extracted from the genome annotation. In addition, we could search the peptide intensities to help selecting AQUA peptide sequences. We mixed the biomass from a nine hours and an 18 hours culture in equal protein amounts. Additionally, we separated the tryptic peptides into 24 fractions using isoelectric focusing, to reduce the peptide complexity. We individually analyzed each fraction by LC-MS/MS and determined 3,265 protein with a FDR <1%. The detected proteins are listed in Table6-2, attached as an electronic *.xlsx file, and below a summarized version. We listed in Table6-2 the *A. gossypii* accession entries, the sequence coverage, the total numbers of identified peptides, numbers of unique peptides, and the determined Exponentially Modified Protein Abundance Index (emPAI) (Ishihama et al. 2005).

Table6-2

accession	percent coverage	number of unique peptides	total number of identified peptides	PAI	emPAI
AAL001W	16.4	5	4	0.80	5.31
AAL002W	24.0	6	9	1.50	30.62
AAL003W	14.0	7	5	0.71	4.18
AAL006C	16.9	7	8	1.14	12.89
AAL007C	50.1	47	97	2.06	114.83
AAL008W	47.3	20	44	2.20	157.49
AAL010W	7.8	6	5	0.83	5.81
AAL011C	33.4	13	21	1.62	40.25
AAL013W	34.7	18	41	2.28	188.57
AAL014C	29.5	8	9	1.13	12.34
AAL016C	12.7	19	19	1.00	9.00
AAL018W	22.3	8	7	0.88	6.50
AAL020C	41.2	17	27	1.59	37.75
AAL021W	47.0	20	27	1.35	21.39
AAL022W	59.3	41	88	2.15	139.07
AAL023W	27.5	11	18	1.64	42.29
AAL025W	61.8	16	25	1.56	35.52
AAL028W	65.5	161	603	3.75	5562.42
AAL029W	24.0	16	30	1.88	73.99
AAL031W	18.7	2	1	0.50	2.16

Appendix 4:

Phospho-proteome analysis of the zero, five, and nine hour time point (Table6-3)

We determined the phospho-proteome by enrichment of phosphorylated peptides using TiO_2 resin (Titansphere TiO_2 5 µm (#5020-75000) in a ratio of 1 mg per 1 mg tryptic peptide mixture. The phospho-peptide enrichment was executed with the zero, five, and nine hour time point. All detected phosphorylation sites are listed in Table6-3, which is attached as an electronic *.xlsx file and as a summary printed below. We sorted Table6-3 based on the *A. gossypii* accession entries. In Table6-3 we listed the identified peptide sequences (column2), the sites of modification (column3), the peptide abundances (column4-6), the abundance changes between the time points (column7 and 9), and the t-test/Anova calculations (column8 and 10).

Table6-3

accession	Sequence	Variable modifications ([position] description)	0h abundance	5h abundance	9h abundance	**Log ratio 0h - 5h**	t-test/Anova 5h-abundance	**Log ratio 0h - 9h**	t-test/Anova 9h-abundance	max (Score)
AAL007C	SGHYTPIPAGASPLKR	[12] Phospho (ST)	26190.65627	286957.429	924226.3771	1.04	2.11E-02	1.55	3.58E-07	31.24
AAL028W	ELGDAYVSSDDEF	[8] Phospho (ST)|[9] Phospho (ST)	1348110.407	2127020.167	7703028.717	0.20	1.58E-01	0.76	4.61E-05	31.00
AAL028W	KELGDAYVSSDDEF	[10] Phospho (ST)	190176.6676	287455.3053	369654.8208	0.18	6.20E-02	0.29	2.85E-02	51.35
AAL028W	KELGDAYVSSDDEF	[9] Phospho (ST)|[10] Phospho (ST)	315507.5686	379922.1188	5321529.044	0.08	8.72E-01	1.23	2.53E-03	28.03
AAL028W	KKELGDAYVSSDDEF	[10] Phospho (ST)|[11] Phospho (ST)	1717.192975	3580.408416	626580.927	0.32	8.67E-01	2.56	1.04E-05	16.89
AAL028W	KKELGDAYVSSDDEF	[11] Phospho (ST)	405033.6626	571275.9568	764641.2864	0.15	1.02E-01	0.28	4.95E-02	39.14
AAL028W	KKELGDAYVSSDDEF	[8] Phospho (Y)|[11] Phospho (ST)	96397.74872	44774.48304	50471.28908	-0.33	3.30E-01	-0.28	3.30E-01	19.71
AAL028W	KKKELGDAYVSSDDEF	[11] Phospho (ST)|[12] Phospho (ST)	64770.2517	73895.67275	992866.3828	0.06	8.84E-01	1.19	1.16E-03	34.10
AAL028W	KKKELGDAYVSSDDEF	[11] Phospho (ST)|[12] Phospho (ST)	9222.78118	33319.62437	1289143.544	0.56	1.36E-01	2.15	2.53E-04	20.40
AAL028W	KKKELGDAYVSSDDEF	[12] Phospho (ST)	30658.58699	13916.53553	73091.20158	-0.34	2.66E-01	0.38	3.34E-01	21.15
AAL028W	LSSAELR	[2] Phospho (ST)	49703.20238	3717.457767	7496.2714	-1.13	1.29E-01	-0.82	9.72E-01	25.73
AAL028W	ELGDAYVSSDDEF	[9] Phospho (ST)	55729.81883	835610.8033	580260.1897	1.18	4.99E-05	1.02	5.84E-07	35.66
AAL046C	SPDAGAADNHKKPR	[1] Phospho (ST)	800.7414462	3442.16982	12.25515859	0.63	9.05E-01	-1.82	3.23E-01	34.54
AAL046C	SPDAGAADNHKKPR	[1] Phospho (ST)	289839.5756	284041.9894	241072.6443	-0.01	6.70E-01	-0.08	4.38E-01	44.35
AAL046C	SPDAGAADNHKKPR	[1] Phospho (ST)	4775.026775	22096.85628	6843.622836	0.67	5.78E-01	0.16	8.08E-01	26.19
AAL072C	HSLGLNEAK	[2] Phospho (ST)	184520.7111	42634.193	28319.87446	-0.64	6.55E-02	-0.81	5.46E-03	16.66
AAL072C	HSLGLNEAKK	[2] Phospho (ST)	41659.78391	10415.19613	2830.477186	-0.60	3.25E-01	-1.17	1.03E-01	17.85
AAL072C	RHSLGLNEAK	[3] Phospho (ST)	16838.2487	88752.92256	29352.46947	0.72	9.23E-02	0.24	2.20E-01	40.56
AAL072C	TGSQSSVYGHAR	[1] Phospho (ST)	79523.20266	54474.15553	52502.13124	-0.16	3.12E-01	-0.18	2.95E-01	45.85
AAL073W	GYQPSPVPTYAAPNEDADPK	[5] Phospho (ST)	1988111.871	729590.1237	306914.5413	-0.44	2.30E-01	-0.81	1.24E-01	41.75

Appendix 5:

Spore proteome based on AUQA peptides or the protein content in a single spore (Table2-20)

Based on two approaches described in Chapter2-C, we applied the All- and Top3-Method to determine protein copy numbers per spore. The top 20 proteins of the first approach based on the percentage in the total defined protein amount is shown in Table2-15, which was extracted from Table2-20. We attached an electronic *.xlsx of Table2-20. The copy numbers per spore (cps) respectively nucleus (cpn) were listed based on both approaches. A summary of Table2-20 is shown below and sorted by the *A. gossypii* accession entries. We listed the cps using the iBAQ method (column3), the T3PQ method (column5), the percentage based on the All-Method (column 4), and the percentage based on the Top3-Method (column6).

Table2-20

accession	common name	cps [iBAQ, regression line]	cps [All-Method; %]	cps [top3, regression line]	cps [Top3-Method; %]
AAL001W	Dom34	469	695	173	980
AAL002W	Rer1	7819	10592	2351	7504
AAL003W	Pgs1	313	470	113	703
AAL004W	Ldb16	12	20	3.8	50
AAL005W	Vma9	3342	4651	348	1688
AAL006C	Stp22	513	757	206	1123
AAL007C	Cbf5	4381	6045	1998	6610
AAL008W	Sis1	7174	9744	1951	6486
AAL009C	Lst8	1593	2269	619	2645
AAL010W	Rfx1	4.5	7.7	0.82	15
AAL011C	Rlp7	453	671	87	573
AAL013W	Nfs1	21740	28513	23206	44837
AAL014C	Pet8	40	64	4.4	56
AAL015W	Dcc1	27	43	5.4	65
AAL016C	Bud3	15	24	14	134
AAL017W	YCL011C-A	97	151	38	301
AAL018W	Hrb1, Gbp2	265	400	82	546
AAL019W	Sgf29	5.0	8.6	0.35	7.7
AAL020C	Mrp7	1131	1628	297	1493
AAL021W	Ilv6	6777	9222	2318	7422

Appendix 6:

Germ bubble proteome (Table3-4; 3-6A/B; 3-7)

The germ bubble proteome and the protein abundances were discussed in Chapter3. First, we determined in Table3-4 the protein copy numbers per nucleus, which is attached as an electronic *.xlsx file. A summary of Table3-4 is printed below which is sorted based on the systematic name of *A. gossypii* genes. The protein abundances determined by the All- and Top3-Methode are listed in column 3 and 4, respectively. We listed the cpn (copies per spore nucleus) using the iBAQ regression line based on AQUA peptides (column5) and the T3PQ regression line based on AQUA peptides (column6). Additionally, we listed the detected peptides, the possible peptides, and the molecular weight.

Next, we determined the protein abundances for germ bubbles by substracting the part corresponding to the non-germinated spores. Short versions of these adjustments are listed in Table3-6A and B of this Appendix. The complete Tables are attached as an electronic *.xlsx file. The lines in the Table are sorted based on the most abundant protein in *A. gossypii* spores. The protein abundances were determined by the All-Method. We listed the protein abundances of spores (column3). The abundances were based on 0.202 mg proteins (column5 in Table3-6A) or on 0.180 mg protein (column5 in Table3-6B) in the germ bubbles of the 5 hours culture. The protein abundances of the spore/germ bubble mixture (column7) was set to the total protein concentration of 0.563 mg and reduced by the spore specific protein content (column5). The germ bubble specific protein content is listed as µg for the summarized table (below).

Table3-4

accession	common name	5h protein abundance (All-Method)	5h protein abundance (Top3-Method)	cpn iBAQ regression line	cpn T3PQ regression line	detected peptides	possible peptides	protein size [MW]
AAL001W	Dom34	49	380	1571	1327	13	31	43576
AAL002W	Rer1	229	847	6354	3274	7	12	22074
AAL003W	Pgs1	14	105	499	312	12	39	57897
AAL004W	Ldb16	0.01	0.1	1	0.1	2	27	35020
AAL005W	Vma9	29	59	994	163	2	4	8287
AAL006C	Stp22	19	161	679	506	6	26	50315
AAL007C	Cbf5	500	2603	12867	11568	37	29	54782
AAL008W	Sis1	264	767	7219	2924	31	19	37744
AAL009C	Lst8	68	473	2127	1700	8	22	33893
AAL010W	Rfx1	1	10	45	23	6	32	92365
AAL011C	Rlp7	113	390	3361	1369	11	16	32527
AAL013W	Nfs1	569	4196	14467	19789	27	29	54384
AAL014C	Pet8	16	69	556	196	7	15	29832
AAL015W	Dcc1	2	18	105	44	7	26	40331
AAL016C	Bud3	0.3	9	16	20	9	95	165582
AAL017W	YCL011C-A	3	29	110	72	1	11	11722
AAL018W	Hrb1, Gbp2	18	107	627	319	7	23	37647
AAL019W	Sgf29	3	11	114	24	5	15	26239
AAL020C	Mrp7	117	471	3458	1689	22	26	40978
AAL021W	Ilv6	417	2057	10913	8877	13	18	32604

Table3-6A

accession	common name	0h protein abundance (All-Method)	0h protein abundance (All-Method)*MW	spore proteins in 5h mix /abMW in 0.202mg	0h spore protein that germinated /abMW in 0.109mg	5h protein abundance (All-Method)	5h protein abundance (All-Method)*MW	total protein in 5h mix /abMW in 0.563mg	bubble specific proteins in µg of the 0.563mg	protein size [MW]
ADL370C	Tef1, Tef2	51734	2580311207	8.76E-03	4.73E-03	59134	2949381078	2.87E-02	19.9850	49876
ABR028C	Cwp1	18731	1313152627	4.46E-03	2.40E-03	19216	1347154543	1.31E-02	8.6716	70107
ABL156C	Fmp45	26259	1235062101	4.19E-03	2.26E-03	22814	1073043051	1.05E-02	6.2654	47034
AER031C	Tdh3	10779	382435010	1.30E-03	7.00E-04	15574	552548454	5.38E-03	4.0867	35479
ABR068C	Fba1	18909	743482636	2.52E-03	1.36E-03	16659	655023655	6.38E-03	3.8600	39319
ABR169W	Atp2	7238	390818072	1.33E-03	7.16E-04	9711	524322798	5.11E-03	3.7832	53992
AFR694W	YLR179C, Tfs1	28817	643566022	2.18E-03	1.18E-03	26686	595973134	5.81E-03	3.6237	22333
ABR222W	Act1	17046	710609117	2.41E-03	1.30E-03	14834	618380449	6.03E-03	3.6145	41688
ABL174C	Ssb2, Ssb1	3401	225126133	7.64E-04	4.12E-04	6458	427452577	4.17E-03	3.4015	66194
AFR512W	Lpd1	8933	476423584	1.62E-03	8.72E-04	9647	514486794	5.01E-03	3.3968	53332
ADL127C	Rpl2A, Rpl2B	6389	174549313	5.92E-04	3.20E-04	14927	407806000	3.97E-03	3.3817	27320
AAL028W	Yef3, Hef3	2136	247233895	8.39E-04	4.53E-04	3610	417894687	4.07E-03	3.2334	115757
AFR142C	Eft2, Eft1	4584	428420837	1.45E-03	7.85E-04	5100	476663584	4.65E-03	3.1912	93469
ACR052W	Sdh1	7509	515931773	1.75E-03	9.45E-04	7348	504875026	4.92E-03	3.1691	68710
AER428W	Om45	13791	742324859	2.52E-03	1.36E-03	10627	572025136	5.57E-03	3.0551	53827
AER187W	Ssa3, Ssa4	4375	307082765	1.04E-03	5.62E-04	5866	411674668	4.01E-03	2.9696	70183
AER085C	Fas1	1488	338739950	1.15E-03	6.20E-04	1672	380551407	3.71E-03	2.5589	227580
AGL272C	Atp1	9161	539458480	1.83E-03	9.88E-04	7569	445731615	4.34E-03	2.5129	58887
ADR206W	Rpl3	2887	126040679	4.28E-04	2.31E-04	6307	275306477	2.68E-03	2.2551	43651
ACL140C	Rps0A, Rps0B	6994	195525615	6.64E-04	3.58E-04	10399	290739251	2.83E-03	2.1697	27957

Table3-6B

accession	common name	0h protein abundance (All-Method)	0h protein abundance (All-Method)*MW	spore proteins in 5h mix /abMW in 0.180mg	0h spore protein that germinated /abMW in 0.130mg	5h protein abundance (All-Method)	5h protein abundance (All-Method)*MW	total protein in 5h mix /abMW in 0.563mg	bubble specific proteins in µg of the 0.563mg	protein size [MW]
ADL370C	Tef1, Tef2	51734	2580311207	7.80E-03	5.64E-03	59134	2949381078	2.87E-02	20.9387	49876
ABR028C	Cwp1	18731	1313152627	3.97E-03	2.87E-03	19216	1347154543	1.31E-02	9.1570	70107
ABL156C	Fmp45	26259	1235062101	3.74E-03	2.70E-03	22814	1073043051	1.05E-02	6.7219	47034
AER031C	Tdh3	10779	382435010	1.16E-03	8.35E-04	15574	552548454	5.38E-03	4.2281	35479
ABR068C	Fba1	18909	743482636	2.25E-03	1.62E-03	16659	655023655	6.38E-03	4.1348	39319
ABR169W	Atp2	7238	390818072	1.18E-03	8.54E-04	9711	524322798	5.11E-03	3.9277	53992
ABR222W	Act1	17046	710609117	2.15E-03	1.55E-03	14834	618380449	6.03E-03	3.8772	41688
AFR694W	YLR179C, Tfs1	28817	643566022	1.95E-03	1.41E-03	26686	595973134	5.81E-03	3.8616	22333
AFR512W	Lpd1	8933	476423584	1.44E-03	1.04E-03	9647	514486794	5.01E-03	3.5729	53332
ABL174C	Ssb2, Ssb1	3401	225126133	6.81E-04	4.92E-04	6458	427452577	4.17E-03	3.4848	66194
ADL127C	Rpl2A, Rpl2B	6389	174549313	5.28E-04	3.81E-04	14927	407806000	3.97E-03	3.4463	27320
ACR052W	Sdh1	7509	515931773	1.56E-03	1.13E-03	7348	504875026	4.92E-03	3.3598	68710
AFR142C	Eft2, Eft1	4584	428420837	1.30E-03	9.36E-04	5100	476663584	4.65E-03	3.3495	93469
AER428W	Om45	13791	742324859	2.24E-03	1.62E-03	10627	572025136	5.57E-03	3.3295	53827
AAL028W	Yef3, Hef3	2136	247233895	7.48E-04	5.40E-04	3610	417894687	4.07E-03	3.3248	115757
AER187W	Ssa3, Ssa4	4375	307082765	9.29E-04	6.71E-04	5866	411674668	4.01E-03	3.0831	70183
AGL272C	Atp1	9161	539458480	1.63E-03	1.18E-03	7569	445731615	4.34E-03	2.7123	58887
AER085C	Fas1	1488	338739950	1.02E-03	7.40E-04	1672	380551407	3.71E-03	2.6841	227580
AFR020W	Ach1	16488	952240834	2.88E-03	2.08E-03	9287	536347842	5.23E-03	2.3470	57754
ADR206W	Rpl3	2887	126040679	3.81E-04	2.75E-04	6307	275306477	2.68E-03	2.3017	43651

To characterize proteins during the development from spores to germ bubbles, we determined the changes of the adjusted protein abundances. Results are listed in Table3-7 attached as an electronic *.xlsx file and a summary version is listed below. We sorted Table3-7 based on the highest ratio from spores to germ bubbles. We listed the protein abundances of zero hours (column2), five hours (column3), and their ratio (column4). For the adjusted germ bubble proteome we determined the ratio (column9) between only germinating spores (column7) and proteins extracted only from germ bubbles (column8).

Table3-7

common name	0h protein abundance (All-Method)	5h protein abundance (All-Method)	**ratio abundance 0h- 5h**	spore proteins in 5h mix /abMW in 0,180mg	total protein in 5h mix /abMW in 0.563mg	0h spore protein that germinated /abMW in 0.130mg	bubble proteins in 5h / in mg	**ratio normalized values 0h - 5h**
Mak3	0.000015	0.16	**10731.55**	9.04E-13	3.13E-08	6.53E-13	3.13E-08	**47880.09**
Spc25	0.00012	0.49	**4057.47**	9.59E-12	1.25E-07	6.92E-12	1.25E-07	**18102.02**
Ftr1	0.0040	7.86	**1984.45**	4.89E-10	3.13E-06	3.53E-10	3.13E-06	**8852.75**
Dsn1	0.000021	0.020	**957.60**	3.69E-12	1.14E-08	2.67E-12	1.14E-08	**4271.17**
Bas1	0.0010	0.65	**664.34**	2.44E-10	5.23E-07	1.77E-10	5.23E-07	**2962.75**
Gdt1	0.0006	0.25	**449.79**	5.12E-11	7.42E-08	3.70E-11	7.42E-08	**2005.47**
Pet309	0.0002	0.077	**328.93**	7.66E-11	8.12E-08	5.53E-11	8.11E-08	**1466.22**
Aur1	0.00010	0.027	**283.10**	1.44E-11	1.31E-08	1.04E-11	1.31E-08	**1261.72**
Sme1	0.0346	9.49	**274.43**	1.15E-09	1.02E-06	8.30E-10	1.01E-06	**1223.07**
Ogg1	0.000056	0.014	**260.25**	7.47E-12	6.26E-09	5.39E-12	6.26E-09	**1159.77**
Esp1	0.0004	0.11	**251.01**	2.41E-10	1.95E-07	1.74E-10	1.95E-07	**1118.57**
YHR132W-	0.0617	14.1	**229.31**	2.52E-09	1.86E-06	1.82E-09	1.86E-06	**1021.75**
Bud14	0.0162	3.66	**226.34**	3.61E-09	2.64E-06	2.61E-09	2.63E-06	**1008.49**
Cwc23	0.0007	0.14	**213.54**	6.34E-11	4.36E-08	4.58E-11	4.36E-08	**951.39**
Snu23	0.0102	2.06	**202.84**	6.80E-10	4.44E-07	4.91E-10	4.44E-07	**903.65**
Pet191	0.0338	6.18	**182.57**	1.27E-09	7.45E-07	9.15E-10	7.44E-07	**813.19**
YMR130W	0.0434	7.78	**179.18**	3.94E-09	2.27E-06	2.84E-09	2.27E-06	**798.05**
Psy2	0.0014	0.24	**171.72**	4.07E-10	2.25E-07	2.94E-10	2.25E-07	**764.81**
Mus81	0.0005	0.080	**158.84**	1.04E-10	5.34E-08	7.53E-11	5.33E-08	**707.30**
YLR407W	0.0039	0.55	**141.56**	2.86E-10	1.30E-07	2.06E-10	1.30E-07	**630.20**

Appendix 7:

Proteome of young mycelia based on AQUA peptides (Table4-5)

We used the iBAQ and the T3PQ method for the use of the AQUA peptides to determined the protein copy numbers per nucleus in young mycelia (Chapter4). Table4-5 shows all determined copy numbers per nucleus and is attached as an electronic *.xlsx file. A section of Table4-5 is shown below and was sorted by the *A. gossypii* accession entries. We listed the cps using the iBAQ method (column5), and the T3PQ method (column6). Additionally, we listed the protein abundances based on the All-Method (column3) and on the Top3-Method (column4). Finally, we included the numbers of detected and possible peptides per protein and the molecular weight of the proteins.

Table4-5

accession	common name	11h protein abundance (All-Method)	11h protein abundance (Top3-Method)	cpn iBAQ [regression line]	cpn T3PQ [regression line]	detected peptides	possible peptides	protein size [MW]
AAL001W	Dom34	20	145	329	183	13	31	43576
AAL002W	Rer1	49	182	801	247	7	12	22074
AAL003W	Pgs1	4	35	63	27	12	39	57897
AAL005W	Vma9	53	107	879	122	2	4	8287
AAL006C	Stp22	1	9	21	4	6	26	50315
AAL007C	Cbf5	1220	5687	19062	24294	37	29	54782
AAL008W	Sis1	643	1986	10142	5978	31	19	37744
AAL009C	Lst8	6	39	103	32	8	22	33893
AAL010W	Rfx1	9	88	149	94	6	32	92365
AAL011C	Rlp7	385	1331	6125	3505	11	16	32527
AAL013W	Nfs1	221	1536	3552	4243	27	29	54384
AAL014C	Pet8	8	42	143	35	7	15	29832
AAL015W	Dcc1	7	50	113	45	7	26	40331
AAL016C	Bud3	4	72	62	72	9	95	165582
AAL017W	YCL011C-A	9	103	159	116	1	11	11722
AAL018W	Hrb1, Gbp2	43	289	705	458	7	23	37647
AAL019W	Sgf29	9	35	146	27	5	15	26239
AAL020C	Mrp7	130	508	2104	970	22	26	40978
AAL021W	Ilv6	248	859	3972	1957	13	18	32604
AAL022W	Idp2, Idp3	573	3481	9068	12625	42	34	46536

Conclusions and outlook

Conclusions and outlook

The system-wide study of protein abundances during the development of *Ashbya gossypii* from spores to young mycelia revile can be used to discover certain protein species involved in cell development. In total, we identified 82% of the ORF-encoding genes at very low FDR of <1% using the target-decoy strategy (Reiter et al., 2009). Interestingly, we could show that *A. gossypii* spores contain already the complete set of proteins for germination. During the development form spores to germ bubbles 800 proteins increase in abundance by de novo protein synthesis and 200 proteins are degraded. A prominent example is the heat shock and chaperon protein Hsp26 (Petko and Lindquist, 1986), the most abundant protein in spores, is fast degraded in germ bubbles and young mycelia. most likely the Hps26 plays an essential role in stabilizing the spore proteome in a non vegetative stage. The development to strictly polarly growing young mycelia was documented using cluster analyses. Specifically, monitoring of the ribosomal proteins reveled a huge increase in the protein abundances. In young mycelia ribosomal proteins are the most abundant protein class. This is in agreement with the fast growth-orientated life style of *A. gossypii*. In contrast to *S. cerevisiae*, *A. gossypii* has an accelerating growth speed (Knechtle et al., 2003). This explains the high demand for protein biosynthesis and the need of high ribosomal amounts. Furthermore, we could show that the intensity-based absolute quantification (iBAQ) technique (Schwanhausser et al.) yields better correlation than the top three protein quantification (T3PQ) method (Silva et al., 2006). In conclusion, we generated a comprehensive quantitative data set that represents a helpful resource for future proteomic studies of *A. gossypii* to further study biological processes of interest in a quantitative manner. Additional, pilot data sets including phospho-peptide analysis were generated during this study, which have great potential to further elucidate molecular mechanisms during growth and development but require further refinements. We used two phospho-peptide specific methods, immobilized metal affinity chelate (IMAC) (Ficarro et al., 2002) and magnetic metal oxide-coated TiO2 chromatographic spheres (Larsen et al., 2005). In total, we generated samples highly enriched for phosphorylated peptides and could identify several hundred unique phospho-sites. However, further optimization of the sample preparation protocol is needed for robust and reproducible enrichment of phospho-peptides and their quantitative analysis.

In the future, we want to focus on comparing the protein expression with mRNA profiling obtained with high-density oligonucleotide microarrays (GeneChips) by Riccarda Rischatsch (Gattiker et al., 2007). Differences can be caused for example by undetected post translational modifications, by translation efficiency, and by protein turnover. In addition, study of evolution by comparing protein abundances with the closely related yeast *S. cerevisiae* (Goffeau et al. 1996; Dietrich et al., 2004) can detect imported changes between both organism. Especially, looking at genes that show high or low sequence alterations can challenge the theory that high abundant proteins are less affected by evolution-based changes than low abundant proteins. We are also interested to identify yet uncharacterized proteins by using a 6 frame shifted protein database. Finally, we are fascinated by continuing experiments based on the SILAC approach (Ong et al., 2002) to study the turnover rate of proteins in *A. gossypii* mycelia by switching cells from labeled to label free media and define the protein changes of isotope to non-labeled peptides by ratio calculations.

References

References

Aebersold, R. and D. R. Goodlett (2001). "Mass spectrometry in proteomics." Chem Rev **101**(2): 269-95.

Aebersold, R. and M. Mann (2003). "Mass spectrometry-based proteomics." Nature **422**(6928): 198-207.

Aebersold, R. H., J. Leavitt, et al. (1987). "Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose." Proc Natl Acad Sci U S A **84**(20): 6970-4.

Ahrens, C. H., E. Brunner, et al. "Quantitative proteomics: a central technology for systems biology." J Proteomics **73**(4): 820-7.

Alberti-Segui, C., F. Dietrich, et al. (2001). "Cytoplasmic dynein is required to oppose the force that moves nuclei towards the hyphal tip in the filamentous ascomycete *Ashbya gossypii*." J Cell Sci **114**(Pt 5): 975-86.

Altmann-Johl, R. and P. Philippsen (1996). "AgTHR4, a new selection marker for transformation of the filamentous fungus *Ashbya gossypii*, maps in a four-gene cluster that is conserved between *A. gossypii* and *Saccharomyces cerevisiae*." Mol Gen Genet **250**(1): 69-80.

Ashby, S. and W. Nowell (1926). "The Fungi of Stigmatomycosis." Ann Bot **40**: 69–84.

Baerenfaller, K., J. Grossmann, et al. (2008). "Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics." Science **320**(5878): 938-41.

Bantscheff, M., M. Schirle, et al. (2007). "Quantitative mass spectrometry in proteomics: a critical review." Anal Bioanal Chem **389**(4): 1017-31.

Barnidge, D. R., E. A. Dratz, et al. (2003). "Absolute quantification of the G protein-coupled receptor rhodopsin by LC/MS/MS using proteolysis product peptides and synthetic peptide standards." Anal Chem **75**(3): 445-51.

Barr, J. R., V. L. Maggio, et al. (1996). "Isotope dilution--mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I." Clin Chem **42**(10): 1676-82.

Batra, L. R. (1973). "Nematosporaceae (Hemiascomycetidae): Taxonomy, pathogenicity, distribution, and vector relations." USDA Technical Bulletin **1469**: 1-71.

Beck, M., A. Schmidt, et al. (2011). "The quantitative proteome of a human cell line." Mol Syst Biol **7**: 549.

Becker, J. and E. A. Craig (1994). "Heat-shock proteins as molecular chaperones." Eur J Biochem **219**(1-2): 11-23.

Bermingham-McDonogh, O., E. B. Gralla, et al. (1988). "The copper, zinc-superoxide dismutase gene of *Saccharomyces cerevisiae*: cloning, sequencing, and biological activity." Proc Natl Acad Sci U S A **85**(13): 4789-93.

Brachat, S. (2003). "Exploring the potential of complete sequence information and synteny: Comparative genome annotation and analysis of *Ashbya gossypii* and *Saccharomyces cerevisiae*." PhD thesis.

Brachat, S., F. S. Dietrich, et al. (2003). "Reinvestigation of the *Saccharomyces cerevisiae* genome annotation by comparison to the genome of a related fungus: *Ashbya gossypii*." Genome Biol **4**(7): R45.

Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." Anal Biochem **72**: 248-54.

Breakspear, A. and M. Momany (2007). "The first fifty microarray studies in filamentous fungi." Microbiology **153**(Pt 1): 7-15.

Brunner, E., C. H. Ahrens, et al. (2007). "A high-quality catalog of the Drosophila melanogaster proteome." Nat Biotechnol **25**(5): 576-83.

Bukau, B. and A. L. Horwich (1998). "The Hsp70 and Hsp60 chaperone machines." Cell **92**(3): 351-66.

Buss, L. J. University of Florida.

Chakraborty, A. and F. E. Regnier (2002). "Global internal standard technology for comparative proteomics." J Chromatogr A **949**(1-2): 173-84.

Chang, E. C., B. F. Crawford, et al. (1991). "Genetic and biochemical characterization of Cu,Zn superoxide dismutase mutants in *Saccharomyces cerevisiae*." J Biol Chem **266**(7): 4417-24.

Cottrell, J. S. (1994). "Protein identification by peptide mass fingerprinting." Pept Res **7**(3): 115-24.

Dammer, K. H. and Grillo-Ravelo (1996). "Verseuchung von Leptoglossus gonagara (Fabr.) mit Nematospora caryli und *Ashbya gossypii* (Ashby et Nowell) Guilliermond in einer Zitrusanlage der Republik Kuba." Arch. Phytopathol. Pflanzenschutz: 26:71.78.

de Sousa Abreu, R., L. O. Penalva, et al. (2009). "Global signatures of protein and mRNA expression levels." Mol Biosyst **5**(12): 1512-26.

Demain, A. L. (1972). "Riboflavin oversynthesis." Annu Rev Microbiol **26**: 369-88.

Dietrich, F. S., S. Voegeli, et al. (2004). "The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome." Science **304**(5668): 304-7.

Domon, B. and R. Aebersold (2006). "Mass spectrometry and protein analysis." Science **312**(5771): 212-7.

Dudoit, S., M. J. van der Laan, et al. (2004). "Multiple testing. Part I. Single-step procedures for control of general type I error rates." Stat Appl Genet Mol Biol **3**: Article13.

Dunkler, A. and J. Wendland (2007). "Use of MET3 promoters for regulated gene expression in *Ashbya gossypii*." Curr Genet **52**(1): 1-10.

Efron, B. and R. Tibshirani (1991). "Statistical data analysis in the computer age." Science **253**(5018): 390-5.

Elias, J. E. and S. P. Gygi (2007). "Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry." Nat Methods **4**(3): 207-14.

Elias, J. E., W. Haas, et al. (2005). "Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations." Nat Methods **2**(9): 667-75.

Fenn, J. B., M. Mann, et al. (1989). "Electrospray ionization for mass spectrometry of large biomolecules." Science **246**(4926): 64-71.

Ficarro, S. B., M. L. McCleland, et al. (2002). "Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*." Nat Biotechnol **20**(3): 301-5.

Field, C. M. and D. Kellogg (1999). "Septins: cytoskeletal polymers or signalling GTPases?" Trends Cell Biol **9**(10): 387-94.

Frazer, H. L. (1944). "Observations on the method of transmission of internal boll disease of cotton by the cotton stainer-bug." Annals of Applied Biology **31**: 271–290.

Gao, J., M. S. Friedrichs, et al. (2005). "Guidelines for the routine application of the peptide hits technique." J Am Soc Mass Spectrom **16**(8): 1231-8.

Gattiker, A., R. Rischatsch, et al. (2007). "Ashbya Genome Database 3.0: a cross-species genome and transcriptome browser for yeast biologists." BMC Genomics **8**: 9.

Gerber, S. A., J. Rush, et al. (2003). "Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS." Proc Natl Acad Sci U S A **100**(12): 6940-5.

Glatter, T., C. Ludwig, et al. (submitted). "Label-free quantitative assessment of different protein digestion protocols reveals that tandem Lys-C/Trypsin proteolysis provides superior cleavage efficiency over standard trypsin digestion."

Goffeau, A., B. G. Barrell, et al. (1996). "Life with 6000 genes." Science **274**(5287): 546, 563-7.

Gordon, J. L., K. P. Byrne, et al. (2009). "Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome." PLoS Genet **5**(5): e1000485.

Grava, S., M. Keller, et al. (2011). "Clustering of nuclei in multinucleated hyphae is prevented by dynein-driven bidirectional nuclear movements and microtubule growth control in *Ashbya gossypii*." Eukaryot Cell **10**(7): 902-15.

Grava, S. and P. Philippsen (2010). "Dynamics of multiple nuclei in *Ashbya gossypii* hyphae depend on the control of cytoplasmic microtubules length by Bik1, Kip2, Kip3, and not on a capture/shrinkage mechanism." Mol Biol Cell **21**(21): 3680-92.

Gygi, S. P., B. Rist, et al. (1999). "Quantitative analysis of complex protein mixtures using isotope-coded affinity tags." Nat Biotechnol **17**(10): 994-9.

Hartwell, L. H. (1971). "Genetic control of the cell division cycle in yeast. IV. Genes controlling bud emergence and cytokinesis." Exp Cell Res **69**(2): 265-76.

Hillenkamp, F., M. Karas, et al. (1991). "Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers." Anal Chem **63**(24): 1193A-1203A.

Hu, Q., R. J. Noll, et al. (2005). "The Orbitrap: a new mass spectrometer." J Mass Spectrom **40**(4): 430-43.

Huang da, W., B. T. Sherman, et al. (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc **4**(1): 44-57.

Ishihama, Y., Y. Oda, et al. (2005). "Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein." Mol Cell Proteomics **4**(9): 1265-72.

Jackson, J. D., V. T. Falciano, et al. (1996). "A likely histone H2A.F/Z variant in *Saccharomyces cerevisiae*." Trends Biochem Sci **21**(12): 466-7.

Johnson, D. I. (1999). "Cdc42: An essential Rho-type GTPase controlling eukaryotic cell polarity." Microbiol Mol Biol Rev **63**(1): 54-105.

Jones, S., M. L. Vignais, et al. (1991). "The CDC25 protein of *Saccharomyces cerevisiae* promotes exchange of guanine nucleotides bound to ras." Mol Cell Biol **11**(5): 2641-6.

Julka, S. and F. Regnier (2004). "Quantification in proteomics through stable isotope coding: a review." J Proteome Res **3**(3): 350-63.

Kapteyn, J. C., H. Van Den Ende, et al. (1999). "The contribution of cell wall proteins to the organization of the yeast cell wall." Biochim Biophys Acta **1426**(2): 373-83.

Kaufmann, A. (2009). "A plasmid collection for PCR-based gene targeting in the filamentous ascomycete *Ashbya gossypii*." Fungal Genet Biol **46**(8): 595-603.

Kaufmann, A. and P. Philippsen (2009). "Of bars and rings: Hof1-dependent cytokinesis in multiseptated hyphae of *Ashbya gossypii*." Mol Cell Biol **29**(3): 771-83.

Keil-Dlouha, V. V., N. Zylber, et al. (1971). "Proteolytic activity of pseudotrypsin." FEBS Lett **16**(4): 291-295.

Kellis, M., B. W. Birren, et al. (2004). "Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*." Nature **428**(6983): 617-24.

Kemper, M., L. Mohlzahn, et al. "A Bnr-like formin links actin to the spindle pole body during sporulation in the filamentous fungus *Ashbya gossypii*." Mol Microbiol **80**(5): 1276-95.

Knechtle, P. (2002). AgSPA2 and AgBOI Control Landmarks of Filamentous Growth in the Filamentous Ascomycete *Ashbya gossypii*.

Philosophisch-Naturwissenschaftlichen Fakultät. Basel, Universität Basel. **Inauguraldissertation**.

Knechtle, P., F. Dietrich, et al. (2003). "Maximal polar growth potential depends on the polarisome component AgSpa2 in the filamentous fungus *Ashbya gossypii*." Mol Biol Cell **14**(10): 4140-54.

Köhli, M., V. Galati, et al. (2008). "Growth-speed-correlated localization of exocyst and polarisome components in growth zones of *Ashbya gossypii* hyphal tips." J Cell Sci **121**(Pt 23): 3878-89.

Kuhn, E., J. Wu, et al. (2004). "Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and 13C-labeled peptide standards." Proteomics **4**(4): 1175-86.

Kurtzman, C. P. (1995). "Relationships among the genera Ashbya, Eremothecium, Holleya and Nematospora determined from rDNA sequence divergence." J Ind Microbiol **14**(6): 523-30.

Kuster, B., M. Schirle, et al. (2005). "Scoring proteomes with proteotypic peptide probes." Nat Rev Mol Cell Biol **6**(7): 577-83.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(5259): 680-5.

Lang, C., S. Grava, et al. (2010). "Mobility, microtubule nucleation and structure of microtubule-organizing centers in multinucleated hyphae of *Ashbya gossypii*." Mol Biol Cell **21**(1): 18-28.

Lange, V., P. Picotti, et al. (2008). "Selected reaction monitoring for quantitative proteomics: a tutorial." Mol Syst Biol **4**: 222.

Larsen, M. R., T. E. Thingholm, et al. (2005). "Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns." Mol Cell Proteomics **4**(7): 873-86.

Li, B., M. Carey, et al. (2007). "The role of chromatin during transcription." Cell **128**(4): 707-19.

Liu, H., R. G. Sadygov, et al. (2004). "A model for random sampling and estimation of relative protein abundance in shotgun proteomics." Anal Chem **76**(14): 4193-201.

Liu, X. F., I. Elashvili, et al. (1992). "Yeast lacking superoxide dismutase. Isolation of genetic suppressors." J Biol Chem **267**(26): 18298-302.

Lopez, N., J. Halladay, et al. (1999). "SSB, encoding a ribosome-associated chaperone, is coordinately regulated with ribosomal protein genes." J Bacteriol **181**(10): 3136-43.

Madden, K. and M. Snyder (1998). "Cell polarity and morphogenesis in budding yeast." Annu Rev Microbiol **52**: 687-744.

Maier, T., M. Guell, et al. (2009). "Correlation of mRNA and protein in complex biological samples." FEBS Lett **583**(24): 3966-73.

Makarov, A. (2000). "Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis." Anal Chem **72**(6): 1156-62.

Mallick, P., M. Schirle, et al. (2007). "Computational prediction of proteotypic peptides for quantitative proteomics." Nat Biotechnol **25**(1): 125-31.

Malmstrom, J., M. Beck, et al. (2009). "Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans." Nature **460**(7256): 762-5.

Meneghini, M. D., M. Wu, et al. (2003). "Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin." Cell **112**(5): 725-36.

Michel, P. E., F. Reymond, et al. (2003). "Protein fractionation in a multicompartment device using Off-Gel isoelectric focusing." Electrophoresis **24**(1-2): 3-11.

Mitchell, P. L. (2004). "Heteroptera as vectors of plant pathogens." Neotrop. Entomol. **33**(5): 519-545.

Moritz, B. and H. E. Meyer (2003). "Approaches for the quantification of protein concentration ratios." Proteomics **3**(11): 2208-20.

Notredame, C., D. G. Higgins, et al. (2000). "T-Coffee: A novel method for fast and accurate multiple sequence alignment." J Mol Biol **302**(1): 205-17.

Ong, S. E., B. Blagoev, et al. (2002). "Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics." Mol Cell Proteomics **1**(5): 376-86.

Pappin, D. J. (1997). "Peptide mass fingerprinting using MALDI-TOF mass spectrometry." Methods Mol Biol **64**: 165-73.

Perkins, D. N., D. J. Pappin, et al. (1999). "Probability-based protein identification by searching sequence databases using mass spectrometry data." Electrophoresis **20**(18): 3551-67.

Petko, L. and S. Lindquist (1986). "Hsp26 is not required for growth at high temperatures, nor for thermotolerance, spore development, or germination." Cell **45**(6): 885-94.

Pfund, C., N. Lopez-Hoyo, et al. (1998). "The molecular chaperone Ssb from *Saccharomyces cerevisiae* is a component of the ribosome-nascent chain complex." Embo J **17**(14): 3981-9.

Picotti, P., B. Bodenmiller, et al. (2009). "Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics." Cell **138**(4): 795-806.

Pokholok, D. K., C. T. Harbison, et al. (2005). "Genome-wide map of nucleosome acetylation and methylation in yeast." Cell **122**(4): 517-27.

Prillinger, H., W. Schweigkofler, et al. (1997). "Phytopathogenic filamentous (*Ashbya, Eremothecium*) and dimorphic fungi (Holleya, Nematospora) with needle-shaped ascospores as new members within the *Saccharomycetaceae.*" Yeast **13**(10): 945-60.

Pruyne, D. and A. Bretscher (2000). "Polarization of cell growth in yeast." J Cell Sci **113 (Pt 4)**: 571-85.

Pruyne, D. and A. Bretscher (2000). "Polarization of cell growth in yeast. I. Establishment and maintenance of polarity states." J Cell Sci **113 (Pt 3)**: 365-75.

Reiter, L., M. Claassen, et al. (2009). "Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry." Mol Cell Proteomics **8**(11): 2405-17.

ResearchInstituteSenckenberg_BotanyDepartment (2008). "West African plants - A Photo Guide." Research Institute Senckenberg Frankfurt/Main, Botany Department.

Richmond, T. J. and C. A. Davey (2003). "The structure of DNA in the nucleosome core." Nature **423**(6936): 145-50.

Rodicio, R., S. Koch, et al. (2006). "KlRHO1 and KlPKC1 are essential for cell integrity signalling in Kluyveromyces lactis." Microbiology **152**(Pt 9): 2635-49.

Roepstorff, P. and J. Fohlman (1984). "Proposal for a common nomenclature for sequence ions in mass spectra of peptides." Biomed Mass Spectrom **11**(11): 601.

Ross, P. L., Y. N. Huang, et al. (2004). "Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents." Mol Cell Proteomics **3**(12): 1154-69.

Sambrook, J. and D. Russel (2001). "Molecular Cloning: A Laboratory Manual."

Sanchez-Marroquin, A., S. Manrique, et al. (1970). "[Production of riboflavin by *Ashbya gossypii* (Ashby and Nowell) Guill. in Agave sp. juice]." Microbiol Esp **23**(2): 127-37.

Schafer, G., E. M. Smith, et al. (2005). "The *Saccharomyces cerevisiae* linker histone Hho1p, with two globular domains, can simultaneously bind to two four-way junction DNA molecules." Biochemistry **44**(50): 16766-75.

Schena, M. (1996). "Genome analysis with gene expression microarrays." Bioessays **18**(5): 427-31.

Schmidt, A., M. Beck, et al. "Absolute quantification of microbial proteomes at different states by directed mass spectrometry." Mol Syst Biol **7**: 510.

Schmidt, A., N. Gehlenborg, et al. (2008). "An integrated, directed mass spectrometric approach for in-depth characterization of complex peptide mixtures." Mol Cell Proteomics **7**(11): 2138-50.

Schmidt, A., J. Kellermann, et al. (2005). "A novel strategy for quantitative proteomics using isotope-coded protein labels." Proteomics **5**(1): 4-15.

Schmitz, H. P., A. Kaufmann, et al. (2006). "From function to shape: a novel role of a formin in morphogenesis of the fungus *Ashbya gossypii*." Mol Biol Cell **17**(1): 130-45.

Schmitz, H. P. and P. Philippsen (2011). "Evolution of multinucleated Ashbya gossypii hyphae from a budding yeast-like ancestor." Fungal Biol **115**(6): 557-68.

Schrimpf, S. P., M. Weiss, et al. (2009). "Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes." PLoS Biol **7**(3): e48.

Schwanhausser, B., D. Busse, et al. "Global quantification of mammalian gene expression control." Nature **473**(7347): 337-42.

Schwartz, J. C., M. W. Senko, et al. (2002). "A two-dimensional quadrupole ion trap mass spectrometer." J Am Soc Mass Spectrom **13**(6): 659-69.

Silva, J. C., M. V. Gorenstein, et al. (2006). "Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition." Mol Cell Proteomics **5**(1): 144-56.

Stahmann, K. P., H. N. Arst, Jr., et al. (2001). "Riboflavin, overproduced during sporulation of *Ashbya gossypii*, protects its hyaline spores against ultraviolet light." Environ Microbiol **3**(9): 545-50.

Stahmann, K. P., J. L. Revuelta, et al. (2000). "Three biotechnical processes using *Ashbya gossypii*, *Candida famata*, or *Bacillus subtilis* compete with chemical riboflavin production." Appl Microbiol Biotechnol **53**(5): 509-16.

Steen, H. and M. Mann (2004). "The ABC's (and XYZ's) of peptide sequencing." Nat Rev Mol Cell Biol **5**(9): 699-711.

Steen, H. and A. Pandey (2002). "Proteomics goes quantitative: measuring protein abundance." Trends Biotechnol **20**(9): 361-4.

Steiner, S., J. Wendland, et al. (1995). "Homologous recombination as the main mechanism for DNA integration and cause of rearrangements in the filamentous ascomycete *Ashbya gossypii*." Genetics **140**(3): 973-87.

Steinman, H. M. (1980). "The amino acid sequence of copper-zinc superoxide dismutase from bakers' yeast." J Biol Chem **255**(14): 6758-65.

Streit, F., V. W. Armstrong, et al. (2002). "Rapid liquid chromatography-tandem mass spectrometry routine method for simultaneous determination of sirolimus, everolimus, tacrolimus, and cyclosporin A in whole blood." Clin Chem **48**(6 Pt 1): 955-8.

Sugimoto, T., A. Morimoto, et al. (2010). "Isolation of an oxalate-resistant *Ashbya gossypii* strain and its improved riboflavin production." J Ind Microbiol Biotechnol **37**(1): 57-64.

Tang, X., J. J. Punch, et al. (2009). "A CAAX motif can compensate for the PH domain of Num1 for cortical dynein attachment." Cell Cycle **8**(19): 3182-90.

Tatusov, R. L., N. D. Fedorova, et al. (2003). "The COG database: an updated version includes eukaryotes." BMC Bioinformatics **4**: 41.

Versele, M. and J. Thorner (2005). "Some assembly required: yeast septins provide the instruction manual." Trends Cell Biol **15**(8): 414-24.

Washburn, M. P., D. Wolters, et al. (2001). "Large-scale analysis of the yeast proteome by multidimensional protein identification technology." Nat Biotechnol **19**(3): 242-7.

Weirich, C. S., J. P. Erzberger, et al. (2008). "The septin family of GTPases: architecture and dynamics." Nat Rev Mol Cell Biol **9**(6): 478-89.

Wendland, J., Y. Ayad-Durieux, et al. (2000). "PCR-based gene targeting in the filamentous fungus *Ashbya gossypii*." Gene **242**(1-2): 381-91.

Wendland, J. and P. Philippsen (2000). "Determination of cell polarity in germinated spores and hyphal tips of the filamentous ascomycete *Ashbya gossypii* requires a rhoGAP homolog." J Cell Sci **113 (Pt 9)**: 1611-21.

Werner-Washburne, M., D. E. Stone, et al. (1987). "Complex interactions among members of an essential subfamily of hsp70 genes in *Saccharomyces cerevisiae*." Mol Cell Biol **7**(7): 2568-77.

White, C. A., N. Oey, et al. (2009). "Global quantitative proteomic profiling through 18O-labeling in combination with MS/MS spectra analysis." J Proteome Res **8**(7): 3653-65.

Wickner, R. B. and M. J. Leibowitz (1976). "Chromosomal genes essential for replication of a double-stranded RNA plasmid of *Saccharomyces cerevisiae*: the killer character of yeast." J Mol Biol **105**(3): 427-43.

Wienkoop, S. and W. Weckwerth (2006). "Relative and absolute quantitative shotgun proteomics: targeting low-abundance proteins in *Arabidopsis thaliana*." J Exp Bot **57**(7): 1529-35.

Wilmen, A. and J. H. Hegemann (1996). "The chromatin of the *Saccharomyces cerevisiae* centromere shows cell-type specific changes." Chromosoma **104**(7): 489-503.

Workman, J. L. and R. E. Kingston (1998). "Alteration of nucleosome structure as a mechanism of transcriptional regulation." Annu Rev Biochem **67**: 545-79.

Wright, M. C. and P. Philippsen (1991). "Replicative transformation of the filamentous fungus *Ashbya gossypii* with plasmids containing *Saccharomyces cerevisiae* ARS elements." Gene **109**(1): 99-105.

Yates, J. R., 3rd, J. K. Eng, et al. (1995). "Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases." Anal Chem **67**(18): 3202-10.

List of abbreviations

List of abbreviations

°C	degree Celsius
µg	microgram
µm	micrometer
A	Alanine
A. gossypii	*Ashbya gossypii*
aa	amino acid
ABC	Ammonium bicarbonate
AFM	Ashbya full medium
Ala	Alanine
APS	ammonium persulfate
AQUA	Absolute QUAntification
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
BCA	bicinchoninic acid
bp	base pair
BSA	Bovine serum albumin
C	Cysteine
COG	Clusters of Orthologous Groups
cpn	copies per nuclei
C-terminus	carboxyl-terminus
Cys	Cysteine
D	Aspartic acid
Da	daltons
DAPI	4',6-diamidino-2-phenylindole
DAVID	Database for Annotation, Visualization and Integrated Discovery
DIC	Differential interference contrast microscopy
DNA	deoxyribonucleic acid
E	Glutamic acid
ESI	electrospray ionization
ESI	electro spray ionization
F	Phenylalanine
FDR	false discovery rate
fg	femtogram
FT-ICR	Fourier Transform Ion Cyclotron Resonance
G	Glycine
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
H	Histidine
h	hour
H2O	water
His	Histidine
HPLC	high-pressure liquid chromatography
I	Isoleucine
iBAQ	intensity based absolute quantification
Ile	Isoleucine
K	Lysine
L	Leucine
LC	liquid chromatography
Leu	Leucine
LIT	linear ion trap
LTQ	linear trap quadrupole
Lys	Lysine
M	Methionine
M	molar concentration
m/z	mass-to-charge
mA	milli ampere
MALDI	Matrix-assisted laser desorption/ionization
Met	Methionine
mg	milligram
min	minutes
ml	milliliter
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MW	molecular weight
N	Asparagine
ng	nanogram
N-terminus	amino-terminus
OGE	OffGel Electrophoresis
ORFs	Open reading frames
P	Proline
PFA	paraformaldehyde
pg	picogram
pH	pondus hydrogenii
Phe	Phenylalanine
pI	Isoelectric point
pKs	acid dissociation constant
PMF	Peptide-Mass-Fingerprinting

Pro	Proline
P-value	probability value
Q	Glutamine
QQQ	triple Quadrupole
Q-TOF	Quadrupole/Time-of-flight
R	Arginine
R2	coefficient of determination
rmp	rounds per minute
RNA	ribonucleic acid
S	Serine
S. cerevisiae	*Saccharomyces cerevisiae*
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser	Serine
SGD	*Saccharomyces cerevisiae* genome datbase
T	Threonine
T3PQ	top 3 protein quantification
TCA	trichloroacetic acid
TECP	Tris (2-carboxyethyl) phosphine
Thr	Threonine
TOF	Time-of-flight
Trp	Tryptophan
Tyr	Tyrosine
Urea	carbonyl diamide
V	Valine
Val	Valine
W	Tryptophan
Y	Tyrosine
µl	microliter

Acknowledgements

Acknowledgements

I would like to thank Peter Philippsen for giving me the opportunity to do my PhD in his laboratory and for all his efforts during the years, especially, for his effort to help me during the project changes and his constant support and generosity. We had a lot of inspiring scientific discussions and he always had good ideas and input for my experimental work. Peter supported me in every way during my thesis and helped me to present my scientific results to the community by attending numerous conferences. I really enjoyed the time I spent in a perfect scientific environment.

I want to express my gratitude to wards Alexander Schmidt for continuously inspiring my enthusiasm for mass spectrometry and teaching me the laboratory and bioinformatics approaches used in this scientific field, and for giving me laboratory and office space after the retirement of Peter. He was stimulating for intellectual exchanges, especially on system-wide proteomics, and provided a nice working atmosphere. Unforgettable are also the pre-work morning ski tours.

Special thanks to Paul Jenö who was extremly supportive during my first attempts to use and understand modern proteomics during my first EU-funded research project.

I want to thank Andreas for supervising me during the first year and for a great time outside the labaratory. Thanks also to Young-Hwa Song for support and teaching biochemistry methods at the start of my PhD. Suzanne Moes for her help and assistance in performing MS experiments during my first PhD project.

Thanks to all former members of the Philippsen group for such enjoyable working atmosphere and thanks to the Proteomics Core Facility for all the fun and exciting seminars.

Thanks to all members of the Marie Curie Research Training Network, Penelope for inspiring meetings and knowledge expanding workshops.

I am thankful to fellow PhD students from other labs in the Biozentrum for being such nice working partners and after-work beers.

Furthermore I want to thank Dr. Hans-Peter Schmitz who introduced me into experiments with *A. gossypii* and supported my move to Basel.

Special thanks to Petra for a great friendship and Mark for discovering with me the North American west coast after the yeast meeting. Thanks to all my friends I could spend time during my PhD doing sports like skiing, beach volleyball and basketball and mountaineering. Also, spending time with all of you at the Rhein during the summer helped me to succeed in my PhD projects.

Outstanding thanks to my whole family for their support over the years and for the relaxing time we spend together. Also for their understanding that I did not had much time for them, because of a typical PhD schedule. Without the constant support of my family I could not have accomplished this goal.

Curriculum vitae

Curriculum vitae

Lars Molzahn

Spalenring 12, CH-4055 Basel, phone: 0041 (0)789 064 927, Lars.Molzahn@gmail.ch

Date of birth:	February 16*th* 1981
Place of birth:	Hamburg
Nationality:	German

Education & Employment History

June 2007 – May 2012	**Applied Microbiology at the Biozentrum, Basel** PhD student “System-wide analysis of absolute protein abundances during the development from spores to multinucleated hyphae of the filamentous fungus *Ashbya gossypii*” - Shot-gun proteome analyses of different development stages - *In vivo* and *In vitro* pull-downs of SH3 domains and proteins
Sep. 2006 – April 2007	**Department of Genetics at the University of Osnabrück** Diploma student “Untersuchungen zur Lokalisierung der Forminproteine *Ag*Bnr1 und *Ag*Bnr2“
July 2002 – August 2006	**Study of Biology at the University of Osnabrück** - Main field of study was genetics, development biology and neurobiology
August 1998 – June 2001	**High school and Abitur**, KBS Moorbeck Norderstedt - Major subjects during the general qualification for the university entrance were nutritional science and biology

Teaching experience

January 2008 – January 2012	**ESBS (Ecole supérieure de biotechnologie Strasbourg) at the Biozentrum, Basel** Tutoring & teaching - Planning and execution of experiments, training of undergrad students and lecturing on Microbiology subjects
May 2004 – July 2005	**Department of Zoology at the University of Osnabrück** Student assistant - Teaching and organize laboratory courses for first- & second-year students

Scholarship

June 2007 – June 2010	**Marie Curie Research Training Network** (RTN) - Funded by the European Commission within the 6th Framework Program

Special skills and expertise

Job based	**Data Reporting and Scientific Writing** Writing Interim reports for the EU commission, monthly presenting project results in group or department seminars, and regularly writing reports on the project progress
Relevant Techniques	LC-MS/MS; DDA (Data-Dependent Acquisition) & MRM (Multiple Reaction Monitoring), Off-Gel electrophoresis (OGE), *In vivo* & *In vitro* pull-downs, Gene targeting, Cloning, Three-dimensional time-lapse (4D) fluorescence microscopy, Immunofluorescence microscopy, practical use of GLP
IT knowledge	**Statistical software:** Spotfire, Excel, Origin **Graphic design:** Adobe Illustrator, Photoshop **Layout – & presentating software:** InDesign, Word, PowerPoint **Mass spectrometry related software:** Progenesis LC-MS, Skyline, Thermo Xcalibur **Microscopy and image processing software:** MetaMorph, ImageJ **Cell biology software:** Clone Manager
Membership	Member of the “Swiss Proteomics Society”
Language skills	German (native) English (Professional working proficiency)

Soft skills

Personal attributes	Excellent teamwork and strong communication skills, self-disciplined, living a culture of mutual respect, conflict mediating person, stable performance under stress
Hobbies	Professional organized ski touring, playing Beach-Volleyball on weekly basis

Qualifications and Courses

April 2009	**Workshop on Light Microscopy, Biozentrum, Basel** / Switzerland Organized by Prof. Dr. Peter Philippsen (Biozentrum) - 4D fluorescence microscopy, phylogenic analyses of yeast species
September 2008	**Workshop on Scientific Writing, CRG, Barcelona** / Spain Organized and realized by Dr. Luis Serrano (Director of the CRG) and Holger Breithaupt (Senior Editor of EMBO reports) - Elements of style: basic rules of clear and precise writing – reporting and live correction

June 2008 **Workshop on Proteomics, AMC, Amsterdam** / Netherlands
Organized by Dave Speijer, Perry Moerland and Ben Distel (Department of Medical Biochemistry & Clinical Proteomics)
- Isoelectric focusing (IEF), In-gel digest, Targetplate spotting and PMF analysis (MALDI TOF), Computer Proteomics/ Ms databases, and introduction in SELDI/Q-STAR/DIGE

April 2008 **Workshop on Structural Biology, EMBL, Hamburg** / Germany
Organized by Dr. Matthias Wilmanns (Head of EMBL Hamburg)
- Affinity chromatography, Protein crystallization, Synchrotron radiation, and protein model building

December 2007 **Workshop on Yeast Cell Biology, EMBL, Heidelberg /** Germany
Organized by Dr. Michael Knop (EMBL) and Prof. Dr. Barbara Winsor (CNRS)
- Yeast 2-hybrid, *In vivo* pull-downs and western blotting

October 2007 **Lab-internship at the EMBL, Hamburg** / Germany
Under supervision of Dr. Young-Hwa Song (Staff Scientist) of the Wilmanns Group
- Cloning, protein expression and purification, CD-spectroscopy and data interpretation

Publications and presentations

Publications

Kemper M, Molzahn L, Lickfeld M, Lang C, Wahlisch S, Schmitz HP: "A Bnr-like formin links actin to the spindle pole body during sporulation in the filamentous fungus *Ashbya gossypii*." Mol Microbiol 2011, 80(5):1276-1295.

Erik Ahrné, Lars Molzahn, Timo Glatter and Alexander Schmidt: "Critical assessment of proteome-wide label-free absolute quantification strategies." Submitted to PROTEOMICS, under revision, April 2013

Oral presentations

Swiss Proteomics Society PhD Student's Symposium: "System-wide proteome analysis during the development of the filamentous fungus *Ashbya gossypii*" December 2012; Biozentrum Basel / Switzerland

SFB (Sonderforschungsbereich) Seminar, invited speaker: "System-wide analysis of absolute protein abundances during the development from spores to multinucleated hyphae of the filamentous fungus *Ashbya gossypii*" May 2012, Uni Osnabrück / Germany

Final Penelope Meeting: "Expression profiling of SH3 proteins in spores, germlings and young mycelium of *Ashbya gossypii"* 2011, CRG, Barcelona / Spain

Third Annual Penelope Meeting: "Expression profiling of SH3 proteins during germination of *Ashbya gossypii*" April 2010, CRG, Barcelona / Spain

Systems biology of filamentous yeast: "Search for SH3-protein interactions using pull-down strategies" June 2009, Uni Osnabrück / Germany

Second Annual Penelope Meeting: "Pull downs of *In vitro* reconstitute SH3 domain interacting complexes" April 2009, Biozentrum Basel / Switzerland

Systems biology of filamentous yeast: "Establishment of interaction partners in yeast networks of Src homology 3 (SH3) proteins" May 2008, Biozentrum Basel / Switzerland

First Annual Penelope Meeting: “Protein interactions of Src homology 3 (SH3) domains in the filamentous fungus *Ashbya gossypii”* April 2008, EMBL, Hamburg / Germany

Posters

Biozentrum symposium: “Expression profiling and absolute quantification of proteins during the germination of the yeast *Ashbya gossypii*” February 2012, Konferenzzentrum St. Chrischona / Switzerland

European Science Foundation Workshop on Affinity Proteomics: “Proteomics of SH3 interaction networks during germination of the filamentous fungus *Ashbya gossypii*” March 2011, Alpbach / Austria

Biozentrum symposium: “Expression profiling of proteins and mRNAs during germination of *Ashbya gossypii*” February 2011, Basel / Switzerland

Yeast Genetics and Molecular Biology Meeting, Genetics Society of America: “Expression profiling of proteins and mRNAs during germination of *Ashbya gossypii*” August 2010, Vancouver / Canada

PhD retreat Biozentrum: “Expression profiling of SH3 proteins during germination of *Ashbya gossypii*” June 2010, Interlaken / Switzerland

Symposium – Molecular Biology of Fungi: “Src Homology 3 (SH3) protein interactions in *Ashbya gossypii*” July 2009, Münster / Germany

PhD retreat Biozentrum: “Evolution of protein interactions between different yeast species based on Src Homology 3 (SH3) domains” June 2009, Engelberg / Switzerland

Symposium – Molecular Biology of Fungi: “A fungal-specific role for a formin protein” July 2007, Hamburg / Germany

Annual conference – Vereinigung für Allgemeine und Angewandte Mikrobiologie: „A fungal-specific role for a formin protein“ 2007, Osnabrück / Germany